Part III

Non-Abelian Gauge Theories

Chapter 14

Invitation: The Parton Model of Hadron Structure

In Part II of this book, we explored the structure of quantum field theories in a formal way. We developed sophisticated calculational algorithms (Chapter 10), derived a formalism for the extraction of scaling laws and asymptotic behavior (Chapter 12), and worked out some of the consequences of spontaneously broken symmetry (Chapter 11). Much of this formalism turned out to have unexpected applications in statistical mechanics. However, we have not yet investigated its implications for elementary particle physics. To do so, we must first ask which particular quantum field theories describe the interactions of elementary particles.

Since the mid-1970s, most high-energy physicists have agreed that the elementary particles that make up matter are a set of fermions, interacting primarily through the exchange of vector bosons. The elementary fermions include the *leptons* (the electron, its heavy counterparts μ and τ, and a neutral, almost massless neutrino corresponding to each of these species), and the *quarks*, whose bound states form the particles with nuclear interactions, mesons and baryons (collectively called hadrons). These fermions interact through three forces: the strong, weak, and electromagnetic interactions. Of these, the *strong interaction* is responsible for nuclear binding and the interactions of the constituents of nuclei, while the *weak interaction* is responsible for radioactive beta decay processes. The *electromagnetic interaction* is the familiar Quantum Electrodynamics, coupled minimally to all charged quarks and leptons. It is not clear that these three forces suffice to explain the most subtle properties of the elementary fermions—we will discuss this question in Chapter 20—but these three forces are certainly the most prominent. All three are now understood to be mediated by the exchange of vector bosons. The equations describing the electromagnetic interaction were discovered by Maxwell, and their quantum mechanical implications have been treated in detail in Part I. The correct theories of the weak and strong interactions were discovered much later.

By the late 1950s, studies of the helicity dependence of weak interaction cross sections and decay rates had shown that the weak interaction involves

a coupling of vector currents built of quark and lepton fields.* It was thus natural to assume that the weak interaction is due to the exchange of very heavy vector bosons, and indeed, such bosons, the W and Z particles, were discovered in experiments at CERN in 1982. But a complete theory of the weak interaction must include not only the correct couplings of the bosons to fermions, but also the equations of motion of the boson fields themselves, the analogue, for the W and Z, of Maxwell's equations. Finding the correct form of these equations was not straightforward, because Maxwell's equations prohibit the generation of a mass for the vector particle. The proper reconciliation of the generalized Maxwell equations with the nonzero W and Z masses turned out to require incorporating into the theory a spontaneously broken symmetry. Chapters 20 and 21 treat this subject in some detail, describing the interplay of vector field theories with spontaneously broken symmetry. This interplay leads to new twists and new phenomena, beyond those discussed in our treatment of spontaneous symmetry breaking in Chapter 11. A complete theory of the weak interaction also requires the simultaneous incorporation of the electromagnetic interaction, forming a unified structure as first hypothesized by Glashow, Weinberg, and Salam.

On the other hand, it was for a long time completely obscure that a theory of exchanged vector bosons could correctly describe the strong interaction. Part of the mystery was that quarks do not exist as isolated species. Their existence, and eventually their quantum numbers, had to be deduced from the spectrum of observable strongly interacting particles. But, in addition, there were complications due to the fact that the strong interactions are strong. The Feynman diagram expansion assumes that the coupling constant is small; when the coupling becomes strong, a large number of diagrams are important (if the series converges at all) and it becomes impossible to pick out the contributions of the elementary interaction vertices. The crucial clue that the strong interactions have a vector character arose from what at first seemed to be just another mystery, the observation that the strong interactions turn themselves off when the momentum transfer is large, in a sense that we will now describe.

Almost Free Partons

In Section 5.1 we computed the cross section for the QED process $e^+e^- \to \mu^+\mu^-$. We then remarked that the corresponding cross section for e^+e^- annihilation into hadrons could be computed in the same way, using a simplistic model in which the quarks are treated as noninteracting fermions. This method gives a surprisingly accurate formula for the cross section, capturing its most important qualitative features. But we deferred the explanation of this puzzle: How can a model of noninteracting quarks represent the behavior of a force that, under other circumstances, is extremely *strong*?

*For an overview of weak interaction phenomenology, see Perkins (1987), Chapter 7, or any other modern particle physics text.

In fact, there are many circumstances in the study of the strong interaction at high energy in which this force has an unexpectedly weak effect. Historically, the first of these appeared in proton-proton collisions. At high energy, above 10 GeV or so in the center of mass, collisions of protons (or any other hadrons) produce large numbers of pions. One might have imagined that these pions would fill all of the allowed phase space, but, in fact, they are mainly produced with momenta almost collinear with the collision axis. The probability of producing a pion with a large component of momentum transverse to the collision axis falls off exponentially in the value of this transverse momentum, suppressing the production substantially for transverse momenta greater than a few hundred MeV.

This phenomenon of limited transverse momentum led to a picture of a hadron as a loosely bound assemblage of many components. In this picture, a proton struck by another proton would be torn into a cloud of pieces. These pieces would have momenta roughly collinear with the original momentum of the proton and would eventually reform into hadrons moving along the collision axis. By hypothesis, these pieces could not absorb a large momentum transfer. We can characterize this hypothesis mathematically as follows: In a high-energy collision, the momenta of the two initial hadrons are almost lightlike. The shattered pieces of the hadrons, arrayed along the collision axis, also have lightlike momenta parallel to the original momentum vectors. This final state can be produced by exchanging momenta q among the pieces in such a way that, though the components of q might be large, the invariant q^2 is always small. The ejection of a hadron at large transverse momentum would require large (spacelike) q^2, but such a process was very rare. Thus it was hypothesized that hadrons were loose clouds of constituents, like jelly, which could not absorb a large q^2.

This picture of hadronic structure was put to a crucial test in the late 1960s, in the SLAC-MIT deep inelastic scattering experiments.† In these experiments, a 20 GeV electron beam was scattered from a hydrogen target, and the scattering rate was measured for large deflection angles, corresponding to large invariant momentum transfers from the electron to a proton in the target. The large momentum transfer was delivered through the electromagnetic rather than the strong interaction, so that the amount of momentum delivered could be computed from the momentum of the scattered electron. In models in which hadrons were complex and softly bound, very low scattering rates were expected.

Instead, the SLAC-MIT experiments saw a substantial rate for hard scattering of electrons from protons. The total reaction rate was comparable to what would have been expected if the proton were an elementary particle scattering according to the simplest expectations from QED. However, only in rare cases did a single proton emerge from the scattering process. The largest part

†For a description of these experiments and their ramifications, see J. I. Friedman, H. W. Kendall, and R. E. Taylor, *Rev. Mod. Phys.* **63**, 573 (1991).

of the rate came from the *deep inelastic* region of phase space, in which the electromagnetic impulse shattered the proton and produced a system with a large number of hadrons.

How could one reconcile the presence of electromagnetic hard scattering processes with the virtual absence of hard scattering in strong interaction processes? To answer this question, Bjorken and Feynman advanced the following simple model, called the *parton model*: Assume that the proton is a loosely bound assemblage of a small number of constituents, called *partons*. These include quarks (and antiquarks), which are fermions carrying electric charge, and possibly other neutral species responsible for their binding. By assumption, these constituents are incapable of exchanging large momenta q^2 through the strong interactions. However, the quarks have the electromagnetic interactions of elementary fermions, so that an electron scattering from a quark can knock it out of the proton. The struck quark then exchanges momentum softly with the remainder of the proton, so that the pieces of the proton materialize as a jet of hadrons. The produced hadrons should be collinear with the direction of the original struck parton.

The parton model, incomplete though it is, imposes a strong constraint on the cross section for deep inelastic electron scattering. To derive this constraint, consider first the cross section for electron scattering from a single constituent quark. We discussed the related process of electron-muon scattering in Section 5.4, and we can borrow that result. Since we imagine the reaction to occur at very high energy, we will ignore all masses. The square of the invariant matrix element in the massless limit is written in a simple form in Eq. (5.71):

$$\frac{1}{4}\sum_{\text{spins}} |\mathcal{M}|^2 = \frac{8e^4Q_i^2}{\hat{t}^2}\left(\frac{\hat{s}^2+\hat{u}^2}{4}\right), \tag{14.1}$$

where $\hat{s}$, $\hat{t}$, $\hat{u}$ are the Mandelstam variables for the electron-quark collision and Q_i is the electric charge of the quark in units of $|e|$. Recall from Eq. (5.73) that, for a collision involving massless particles, $\hat{s}+\hat{t}+\hat{u}=0$. Then the differential cross section in the center of mass system is

$$\begin{aligned}\frac{d\sigma}{d\cos\theta_{\text{CM}}} &= \frac{1}{2\hat{s}}\frac{1}{16\pi}\frac{8e^4Q_i^2}{\hat{t}^2}\left(\frac{\hat{s}^2+\hat{u}^2}{4}\right)\\ &= \frac{\pi\alpha^2Q_i^2}{\hat{s}}\left(\frac{\hat{s}^2+(\hat{s}+\hat{t})^2}{\hat{t}^2}\right).\end{aligned} \tag{14.2}$$

Or, since $\hat{t} = -\hat{s}(1-\cos\theta_{\text{CM}})/2$,

$$\frac{d\sigma}{d\hat{t}} = \frac{2\pi\alpha^2Q_i^2}{\hat{s}^2}\left(\frac{\hat{s}^2+(\hat{s}+\hat{t})^2}{\hat{t}^2}\right). \tag{14.3}$$

To make use of this result, we must relate the invariants $\hat{s}$ and $\hat{t}$ to experimental observables of electron-proton inelastic scattering. The kinematic variables are shown in Fig. 14.1. The momentum transfer q from the electron

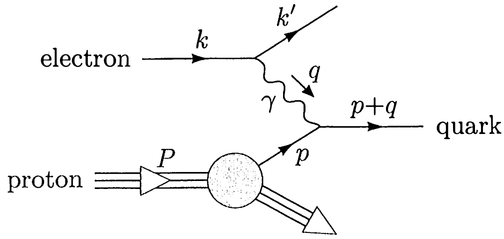


Figure 14.1. Kinematics of deep inelastic electron scattering in the parton model.

can be measured by measuring the final momentum and energy of the electron, without using any information from the hadronic products. Since q^μ is a spacelike vector, one conventionally expresses its invariant square in terms of a positive quantity Q, with

$$Q^2 \equiv -q^2. \tag{14.4}$$

Then the invariant $\hat{t}$ is simply $-Q^2$.

Expressing $\hat{s}$ in terms of measurable quantities is more difficult. If the collision is viewed from the electron-proton center of mass frame, and we visualize the proton as a loosely bound collection of partons (and continue to ignore masses), we can characterize a given parton by the fraction of the proton's total momentum that it carries. We denote this *longitudinal fraction* by the parameter ξ, with $0 < \xi < 1$. For each species i of parton, for example, up-type quarks with electric charge $Q_i = +2/3$, there will be a function $f_i(\xi)$ that expresses the probability that the proton contains a parton of type i and longitudinal fraction ξ. The expression for the total cross section for electron-proton inelastic scattering will contain an integral over the value of ξ for the struck parton. The momentum vector of the parton is then $p = \xi P$, where P is the total momentum of the proton. Thus, if k is the initial electron momentum,

$$\hat{s} = (p+k)^2 = 2p \cdot k = 2\xi P \cdot k = \xi s, \tag{14.5}$$

where s is the square of the electron-proton center of mass energy.

Remarkably, ξ can also be determined from measurements of only the electron momentum, if one makes the assumption that the electron-parton scattering is elastic. Since the scattered parton has a mass small compared to s and Q^2,

$$0 \approx (p+q)^2 = 2p \cdot q + q^2 = 2\xi P \cdot q - Q^2. \tag{14.6}$$

Thus

$$\xi = x, \qquad \text{where} \quad x \equiv \frac{Q^2}{2P \cdot q}. \tag{14.7}$$

From each scattered electron, one can determine the values of Q^2 and x for the scattering process. The parton model then predicts the event distribution

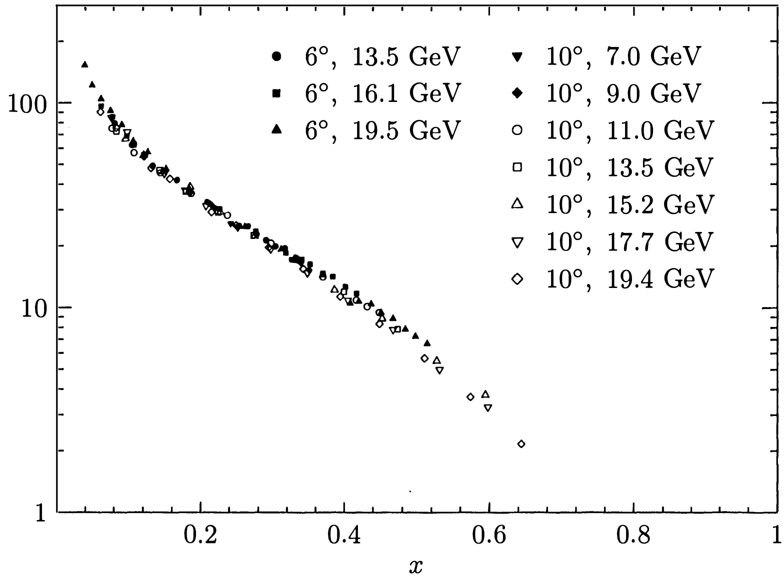


Figure 14.2. Test of Bjorken scaling using the $e^- p$ deep inelastic scattering cross sections measured by the SLAC-MIT experiment, J. S. Poucher, et. al., *Phys. Rev. Lett.* **32**, 118 (1974). We plot $d^2\sigma/dxdQ^2$ divided by the factor (14.9) against x, for the various initial electron energies and scattering angles indicated. The data span the range 1 GeV^2 $< Q^2 <$ 8 GeV^2.

in the x-Q^2 plane. Using the parton distribution functions $f_i(\xi)$, evaluated at $\xi = x$, and the cross-section formula (14.3), we find the distribution

$$\frac{d^2\sigma}{dxdQ^2} = \sum_i f_i(x)Q_i^2 \cdot \frac{2\pi\alpha^2}{Q^4}\left[1 + \left(1 - \frac{Q^2}{xs}\right)^2\right]. \tag{14.8}$$

The distribution functions $f_i(x)$ depend on the details of the structure of the proton and it is not known how to compute them from first principles. But formula (14.8) still makes a striking prediction, that the deep inelastic scattering cross section, when divided by the factor

$$\frac{1 + (1 - Q^2/xs)^2}{Q^4} \tag{14.9}$$

to remove the kinematic dependence of the QED cross section, gives a quantity that depends only on x and is independent of Q^2. This behavior is known as *Bjorken scaling.* Indeed, the data from the SLAC-MIT experiment exhibited Bjorken scaling to about 10% accuracy for values of Q above 1 GeV, as shown in Fig. 14.2.

Bjorken scaling is, essentially, the statement that the structure of the proton looks the same to an electromagnetic probe no matter how hard the proton is struck. In the frame of the proton, the energy of the exchanged

virtual photon is

$$q^0 = \frac{P \cdot q}{m} = \frac{Q^2}{2xm}, \tag{14.10}$$

where m is the proton mass. The reciprocal of this energy transfer is, roughly, the duration of the scattering process as seen by the components of the proton. This time should be compared to the reciprocal of the proton mass, which is the characteristic time over which the partons interact. The deep inelastic regime occurs when $q^0 \gg m$, that is, when the scattering is very rapid compared to the normal time scales of the proton. Bjorken scaling implies that, during such a rapid scattering process, interactions among the constituents of the proton can be ignored. We might imagine that the partons are approximately free particles over the very short times scales corresponding to energy transfers of a GeV or more, though they have strong interactions on longer time scales.

Asymptotically Free Partons

The picture of the proton structure implied by Bjorken scaling was beautifully simple, but it raised new, fundamental questions. In quantum field theory, fermions interact by exchanging virtual particles. These virtual particles can have arbitrarily high momenta, hence the fluctuations associated with them can occur on arbitrarily short time scales. Quantum field theory processes do not turn themselves off at short times to reveal free-particle equations. Thus the discovery of Bjorken scaling suggested a conflict between the observation of almost free partons and the basic principles of quantum field theory.

The resolution of this paradox came from the renormalization group. In Chapter 12 we saw that coupling constants vary with distance scale. In QED and ϕ^4 theory, we found that the couplings become strong at large momenta and weak at small momenta. However, we noted the possibility that, in some theories, the coupling constant could have the opposite behavior, becoming strong at small momenta or large times but weak at large momenta or short times. We referred to such behavior as *asymptotic freedom.* Section 13.3 discussed an example of an asymptotically free quantum field theory, the nonlinear sigma model in two dimensions. The problem posed in the previous paragraph would be resolved if there existed a suitable asymptotically free quantum field theory in four dimensions that could describe the interaction and binding of quarks. Then, at least to some level of approximation, the strong interaction described by this theory would turn off in large-momentum-transfer or short-time processes.

At the time of the discovery of Bjorken scaling, no asymptotically free field theories in four dimensions were known. Then, in the early 1970s, 't Hooft, Politzer, Gross, and Wilczek discovered a class of such theories. These are the *non-Abelian gauge theories*: theories of interacting vector bosons that can be constructed as generalizations of quantum electrodynamics. It was subsequently shown that these are the only asymptotically free field theories

in four dimensions. This discovery gave the crucial clue for the construction of the fundamental theory of the strong interactions. Apparently, the quarks are bound together by interacting vector bosons (called *gluons*) of precisely this type.

However, these gauge theories cannot precisely reproduce the expectations of strict Bjorken scaling. The differences between the free parton model and the quantum field theory model with asymptotic freedom appear when one moves to a higher level of accuracy in measurements of deep inelastic scattering and other strong interaction processes involving large momentum transfer. In an asymptotically free quantum field theory, the coupling constant is still nonzero at any finite momentum transfer. In fact, the final evolution of the coupling to zero is very slow, logarithmic in momentum. Thus, at some level, one must find small corrections to Bjorken scaling, associated with the exchange or emission of high-momentum gluons. Similarly, the other qualitative simplifications of hadron physics at high momentum transfer—for example, the phenomenon of limited transverse momentum in hadron-hadron collisions—should be only approximate, receiving corrections due to gluon exchange and emission. Thus the predictions of an asymptotically free theory of the strong interaction are twofold. On one hand, such a theory predicts qualitative simplifications of behavior at high momentum. But, on the other hand, such a theory predicts a specific pattern of corrections to this behavior.

In fact, particle physics experiments of the 1970s revealed precisely this picture. Bjorken scaling was found to be only an approximate relation, showing violations that correspond to a slow evolution of the parton distributions $f_i(x)$ over a logarithmic scale in Q^2. The rate of particle production in hadron-hadron collisions was found to decrease only as a power rather than exponentially at very large values of the transverse momentum, and the particles produced at large transverse momentum were shown to be associated with jets of hadrons created by the soft evolution of a hard-scattered quark or gluon. Most remarkably, the forms of the cross sections found for these and other deviations from scaling did, finally, give direct evidence for the vector character of the elementary field that mediates the strong interaction.

We will review all of these phenomena in Chapter 17, as we study the particular gauge theory that describes the strong interactions. First, however, we must learn how to construct non-Abelian gauge theories and how to work out their predictions using Feynman diagrams. Throughout our analysis of these theories, the renormalization group will play an essential role. One of the very beautiful aspects of the study of non-Abelian gauge theories is the way in which the most powerful general ideas of quantum field theory acquire even more strength as they intertwine with the specific features of these particular, intricately built models. This interplay between general principles and the specific features of gauge theories will be the major theme of Part III of this book.

Chapter 15

Non-Abelian Gauge Invariance

So far in this book we have worked with a rather limited class of quantum fields and interactions, restricting our attention to scalar field theories, Yukawa theory, and Quantum Electrodynamics. It is hardly surprising that these theories are not sufficient to describe all of the known interactions of elementary particles. But what other theories are possible, given that the Lagrangian of a renormalizable theory can contain no terms of mass dimension higher than 4?

The most natural theories to try next would be ones with interactions among vector fields, of the form $A^\mu A^\nu \partial_\mu A_\nu$ or A^4. Sensible theories of this type are difficult to construct, however, because of the negative-norm states produced by the time component A^0 of the vector field operator. In Section 5.5 we saw that these negative-norm states cause no difficulty in QED: They are effectively canceled out by the longitudinal polarization states, by virtue of the Ward identity. The Ward identity, in turn, follows from the invariance of the QED Lagrangian under local gauge transformations. Perhaps, then, if we can generalize the principle of local gauge invariance, it will lead us to the construction of other sensible theories of vector particles.

The goal of this chapter is to do just that. First we will return briefly to the study of QED, this time taking the gauge symmetry to be fundamental and deriving the rest of the theory from this principle. Then, in Section 15.2, we will see that the gauge invariance of electrodynamics is only the most trivial example of an infinite-parameter symmetry, and that the more general examples lead to other interesting Lagrangians. These field theories, the first of which was constructed by Yang and Mills,* generalize electrodynamics in a profound way. They are theories of multiple vector particles, whose interactions are strongly constrained by the symmetry principle. In subsequent chapters we will study the quantization of these theories and their application to the real world of elementary particle physics.

*C. N. Yang and R. Mills, *Phys. Rev.* **96**, 191 (1954).

15.1 The Geometry of Gauge Invariance

In Section 4.1 we wrote down the Lagrangian of Quantum Electrodynamics and noted the curious fact that it is invariant under a very large group of transformations (4.6), allowing an independent symmetry transformation at every point in spacetime. This invariance is the famous *gauge symmetry* of QED. From the modern viewpoint, however, gauge symmetry is not an incidental curiosity, but rather the fundamental principle that determines the form of the Lagrangian. Let us now review the elements of the theory, taking the modern viewpoint.

We begin with the complex-valued Dirac field $\psi(x)$, and stipulate that our theory should be invariant under the transformation

$$\psi(x) \to e^{i\alpha(x)}\psi(x). \tag{15.1}$$

This is a phase rotation through an angle $\alpha(x)$ that varies arbitrarily from point to point. How can we write a Lagrangian that is invariant under this transformation? As long as we consider terms in the Lagrangian that have no derivatives, this is easy: We simply write the same terms that are invariant to global phase rotations. For example, the fermion mass term

$$m\bar{\psi}\psi(x)$$

is permitted by global phase invariance, and the local invariance gives no further restriction.

The difficulty arises when we try to write terms including derivatives. The derivative of $\psi(x)$ in the direction of the vector n^μ is defined by the limiting procedure

$$n^\mu \partial_\mu \psi = \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left[\psi(x+\epsilon n) - \psi(x)\right]. \tag{15.2}$$

However, in a theory with local phase invariance, this definition is not very sensible, since the two fields that are subtracted, $\psi(x+\epsilon n)$ and $\psi(x)$, have completely different transformations under the symmetry (15.1). The quantity $\partial_\mu\psi$, in other words, has no simple tranformation law and no useful geometric interpretation.

In order to subtract the values of $\psi(x)$ at neighboring points in a meaningful way, we must introduce a factor that compensates for the difference in phase transformations from one point to the next. The simplest way to do this is to define a scalar quantity $U(y,x)$ that depends on the two points and has the transformation law

$$U(y,x) \to e^{i\alpha(y)}U(y,x)e^{-i\alpha(x)} \tag{15.3}$$

simultaneously with (15.1). At zero separation, we set $U(y,y)=1$; in general, we can require $U(y,x)$ to be a pure phase: $U(y,x) = \exp[i\phi(y,x)]$. With this definition, the objects $\psi(y)$ and $U(y,x)\psi(x)$ have the same transformation law, and we can subtract them in a manner that is meaningful despite the

local symmetry. Thus we can define a sensible derivative, called the *covariant derivative*, as follows:

$$n^\mu D_\mu \psi = \lim_{\epsilon \to 0} \frac{1}{\epsilon}\left[\psi(x+\epsilon n) - U(x+\epsilon n, x)\psi(x)\right]. \tag{15.4}$$

To make this definition explicit, we need an expression for the comparator $U(y,x)$ at infinitesimally separated points. If the phase of $U(y,x)$ is a continuous function of the positions y and x, then $U(y,x)$ can be expanded in the separation of the two points:

$$U(x+\epsilon n, x) = 1 - ie\,\epsilon n^\mu A_\mu(x) + \mathcal{O}(\epsilon^2). \tag{15.5}$$

Here we have arbitrarily extracted a constant e. The coefficient of the displacement ϵn^μ is a new vector field $A_\mu(x)$. Such a field, which appears as the infinitesimal limit of a comparator of local symmetry transformations, is called a *connection*. The covariant derivative then takes the form

$$D_\mu \psi(x) = \partial_\mu \psi(x) + ieA_\mu \psi(x). \tag{15.6}$$

By inserting (15.5) into (15.3), one finds that A_μ transforms under this local gauge transformation as

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu \alpha(x). \tag{15.7}$$

To check that all of these expressions are consistent, we can transform $D_\mu\psi(x)$ according to Eqs. (15.1) and (15.7):

$$\begin{aligned} D_\mu \psi(x) &\to \left[\partial_\mu + ie\big(A_\mu - \frac{1}{e}\partial_\mu \alpha\big)\right] e^{i\alpha(x)}\psi(x) \\ &= e^{i\alpha(x)}\big(\partial_\mu + ieA_\mu\big)\psi(x) = e^{i\alpha(x)} D_\mu \psi(x). \end{aligned} \tag{15.8}$$

Thus the covariant derivative transforms in the same way as the field ψ, exactly as we constructed it to in the original definition (15.4).

We have now recovered most of the familiar ingredients of the QED Lagrangian. From our current viewpoint, however, the definition of the covariant derivative and the transformation law for the connection A_μ follow from the postulate of local phase rotation symmetry. Even the very existence of the vector field A_μ is a consequence of local symmetry: Without it we could not write an invariant Lagrangian involving derivatives of ψ.

More generally, our present analysis gives us a way to construct all possible Lagrangians that are invariant under the local symmetry. In any term with derivatives of ψ, replace these with covariant derivatives. According to Eq. (15.8), these transform in exactly the same manner as ψ itself. Therefore any combination of ψ and its covariant derivatives that is invariant under a global phase rotation (and only these combinations) will also be locally invariant.

To complete the construction of a locally invariant Lagrangian, we must find a kinetic energy term for the field A_μ: a locally invariant term that depends on A_μ and its derivatives, but not on ψ. This term can be constructed

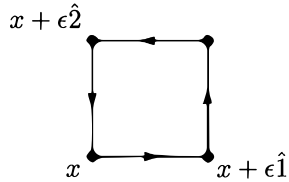


Figure 15.1. Construction of the field strength by comparisons around a small square in the (1, 2) plane.

either integrally, from the comparator $U(y,x)$, or infinitesimally, from the covariant derivative.

Working from $U(y,x)$, we will need to extend our explicit formula (15.5) to the next term in the expansion in ϵ. Using the assumption that $U(y,x)$ is a pure phase and the restriction $(U(x,y))^\dagger = U(y,x)$, it follows that

$$U(x+\epsilon n, x) = \exp\left[-ie\epsilon n^\mu A_\mu(x+\tfrac{\epsilon}{2}n) + \mathcal{O}(\epsilon^3)\right]. \tag{15.9}$$

(Relaxing these restrictions introduces additional vector fields into the theory; this is an unnecessary complication.) Using this expansion for $U(y,x)$, we link together comparisons of the phase direction around a small square in spacetime. For definiteness, we take this square to lie in the $(1,2)$-plane, as defined by the unit vectors $\hat{1}$, $\hat{2}$ (see Fig. 15.1). Define $\mathbf{U}(x)$ to be the product of the four comparisons around the corners of the loop:

$$\begin{aligned}\mathbf{U}(x) \equiv U(x, x+\epsilon\hat{2})&U(x+\epsilon\hat{2}, x+\epsilon\hat{1}+\epsilon\hat{2})\\ &\times U(x+\epsilon\hat{1}+\epsilon\hat{2}, x+\epsilon\hat{1})U(x+\epsilon\hat{1}, x).\end{aligned} \tag{15.10}$$

The transformation law (15.3) for U implies that $\mathbf{U}(x)$ is locally invariant. In the limit $\epsilon \to 0$, it will therefore give us a locally invariant function of A_μ. To find the form of this function, insert the expansion (15.9) to obtain

$$\begin{aligned}\mathbf{U}(x) = \exp\Big\{ -i\epsilon e\big[&-A_2(x+\tfrac{\epsilon}{2}\hat{2}) - A_1(x+\tfrac{\epsilon}{2}\hat{1}+\epsilon\hat{2})\\ &+ A_2(x+\epsilon\hat{1}+\tfrac{\epsilon}{2}\hat{2}) + A_1(x+\tfrac{\epsilon}{2}\hat{1})\big] + \mathcal{O}(\epsilon^3)\Big\}.\end{aligned} \tag{15.11}$$

When we expand the exponent in powers of ϵ, this reduces to

$$\mathbf{U}(x) = 1 - i\epsilon^2 e\big[\partial_1 A_2(x) - \partial_2 A_1(x)\big] + \mathcal{O}(\epsilon^3). \tag{15.12}$$

Therefore the structure

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{15.13}$$

is locally invariant. Of course, $F_{\mu\nu}$ is the familiar electromagnetic field tensor, and its invariance under (15.7) can be checked directly. The preceding construction, however, shows us the geometrical origin of the structure of $F_{\mu\nu}$. Any function that depends on A_μ only through $F_{\mu\nu}$ and its derivatives is locally invariant. More general functions, such as the vector field mass term

$A_\mu A^\mu$, transform under (15.7) in ways that cannot be compensated and thus cannot appear in an invariant Lagrangian.

A related argument for the invariance of $F_{\mu\nu}$ can be made using the covariant derivative. We have seen above that, if a field has the local transfomation law (15.1), then its covariant derivative has the same transformation law. Thus the second covariant derivative of ψ also transforms according to (15.1). The same conclusion holds for the commutator of covariant derivatives:

$$[D_\mu, D_\nu]\psi(x) \to e^{i\alpha(x)}[D_\mu, D_\nu]\psi(x). \tag{15.14}$$

However, the commutator is not itself a derivative at all:

$$\begin{aligned}[D_\mu, D_\nu]\psi &= [\partial_\mu, \partial_\nu]\psi + ie\big([\partial_\mu, A_\nu] - [\partial_\nu, A_\mu]\big)\psi - e^2[A_\mu, A_\nu]\psi \\ &= ie\big(\partial_\mu A_\nu - \partial_\nu A_\mu\big)\cdot\psi.\end{aligned} \tag{15.15}$$

That is,

$$[D_\mu, D_\nu] = ieF_{\mu\nu}. \tag{15.16}$$

On the right-hand side of (15.14), the factor $\psi(x)$ accounts for the entire transformation law, so the multiplicative factor $F_{\mu\nu}$ must be invariant. One can visualize the commutator of covariant derivatives as the comparison of comparisons across a small square; fundamentally, therefore, this argument is equivalent to that of the previous paragraph.

Whatever the method of proving the invariance of $F_{\mu\nu}$, we have now assembled all of the ingredients we need to write the most general locally invariant Lagrangian for the electron field ψ and its associated connection A_μ. This Lagrangian must be a function of ψ and its covariant derivatives, and of $F_{\mu\nu}$ and its derivatives, and must be invariant to global phase transformations. Up to operators of dimension 4, there are only four possible terms:

$$\mathcal{L}_4 = \bar{\psi}(i\not{D})\psi - \frac{1}{4}(F_{\mu\nu})^2 - c\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu} - m\bar{\psi}\psi. \tag{15.17}$$

By adjusting the normalization of the fields ψ and A_μ, we have set the coefficients of the first two terms to their standard values. This normalization of A_μ requires the arbitrary scale factor e in our original definition (15.5) of A_μ. The third term violates the discrete symmetries P and T, so we may exclude it if we postulate these symmetries.† Then $\mathcal{L}_4$ contains only two free parameters, the scale factor e and the coefficient m.

By using operators of dimension 5 and 6, we can form many additional gauge-invariant combinations:

$$\mathcal{L}_6 = ic_1\bar{\psi}\sigma_{\mu\nu}F^{\mu\nu}\psi + c_2(\bar{\psi}\psi)^2 + c_3(\bar{\psi}\gamma^5\psi)^2 + \cdots. \tag{15.18}$$

More allowed terms appear at each higher order in mass dimension. But all of these terms are nonrenormalizable interactions. In the language of Section 12.1, they are irrelevant to physics in four dimensions in the limit where the cutoff is taken to infinity.

---

†The general systematics of P, C, and T violation are discussed in Section 20.3.

We have now reached a remarkable conclusion. We began by postulating that the electron field obeys the local symmetry (15.1). From this postulate, we showed that there must be an electromagnetic vector potential. Further, the symmetry principle implies that the most general Lagrangian in four dimensions that is renormalizable (or relevant, in Wilson's sense) is the general form $\mathcal{L}_4$. If we insist that this Lagrangian also be invariant under time reversal or parity, we are led uniquely to the Maxwell-Dirac Lagrangian that is the basis of quantum electrodynamics.

15.2 The Yang-Mills Lagrangian

If the simple geometrical constructions of the previous section yield Maxwell's theory of electrodynamics, then surely it must be possible to construct other interesting theories by starting with more general geometrical principles. Yang and Mills proposed that the argument of the previous section could be generalized from local phase rotation invariance to invariance under any continuous symmetry group. In this section, we will introduce this generalization of local symmetry. For most of the discussion, we will consider our local symmetry to be the three-dimensional rotation group, $O(3)$ or $SU(2)$, since in this case the necessary group theory should be familiar. At the end of the section, we will generalize further to the case of an arbitrary local symmetry.

Consider, then, the following generalization of the phase rotation (15.1): Instead of a single fermion field, we start with a doublet of Dirac fields,

$$\psi = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}, \tag{15.19}$$

which transform into one another under abstract three-dimensional rotations as a two-component spinor:

$$\psi \to \exp\left(i\alpha^i \frac{\sigma^i}{2}\right)\psi. \tag{15.20}$$

Here σ^i are the Pauli sigma matrices, and, as usual, a sum over repeated indices is implied. It is important to distinguish this abstract transformation from a rotation in physical three-dimensional space; in their original paper, Yang and Mills considered (ψ_1, ψ_2) to be the proton-neutron doublet as it is transformed under isotopic spin. As in the case of a phase rotation, it is not hard to construct Lagrangians for ψ that are invariant to (15.20) as a global symmetry.

We now promote (15.20) to a local symmetry, by insisting that the Lagrangian be invariant to this transformation for α^i an arbitrary function of x. Write this transformation as

$$\psi(x) \to V(x)\psi(x), \qquad \text{where } V(x) = \exp\left(i\alpha^i(x)\frac{\sigma^i}{2}\right). \tag{15.21}$$

We can construct a suitable Lagrangian by applying the methods of the previous section. However, we will encounter a number of additional complications, due to the fact that there are now three orthogonal symmetry motions, which do not commute with one another. This feature is sufficiently important to earn a special name for theories that have it: We refer to the *Abelian* symmetry group of electrodynamics, and the *non-Abelian* symmetry group of the more general theories. The field theory associated with a noncommuting local symmetry is termed a *non-Abelian gauge theory.*

To construct a Lagrangian that is invariant under this new group of transformations, we must again define a covariant derivative that transforms in a simple way. Again we use the definition (15.4), but since ψ is now a two-component object, the comparator $U(y,x)$ must be a 2×2 matrix. The transformation law for $U(y,x)$ is now

$$U(y,x) \to V(y)\,U(y,x)\,V^\dagger(x), \tag{15.22}$$

where $V(x)$ is as in (15.21), and again we set $U(y,y) = 1$. At points $x \neq y$ we can consistently restrict $U(y,x)$ to be a unitary matrix. Near $U = 1$, any such matrix can be expanded in terms of the Hermitian generators of $SU(2)$; thus for infinitesimal separation we can write

$$U(x+\epsilon n, x) = 1 + ig\epsilon n^\mu A^i_\mu \frac{\sigma^i}{2} + \mathcal{O}(\epsilon^2). \tag{15.23}$$

Here g is a constant, extracted for later convenience. Inserting this expansion into the definition (15.4) of the covariant derivative, we find the following expression for the covariant derivative associated with local $SU(2)$ symmetry:

$$D_\mu = \partial_\mu - igA^i_\mu \frac{\sigma^i}{2}. \tag{15.24}$$

This covariant derivative requires three vector fields, one for each generator of the transformation group.

We can find the gauge transformation law of the connection A^i_μ by inserting the expansion (15.23) into the transformation law (15.22):

$$1 + ig\epsilon n^\mu A^i_\mu \frac{\sigma^i}{2} \to V(x+\epsilon n)\Big(1 + ig\epsilon n^\mu A^i_\mu \frac{\sigma^i}{2}\Big)V^\dagger(x). \tag{15.25}$$

We must expand the right-hand side to order ϵ, taking care that the various Pauli matrices do not commute with one another. The expansion of $V(x+\epsilon n)$ is conveniently done using the identity

$$\begin{aligned} V(x+\epsilon n)V^\dagger(x) &= \Big[\Big(1 + \epsilon n^\mu \frac{\partial}{\partial x^\mu} + \mathcal{O}(\epsilon^2)\Big)V(x)\Big]V^\dagger(x) \\ &= 1 + \epsilon n^\mu \Big(\frac{\partial}{\partial x^\mu} V(x)\Big)V^\dagger(x) + \mathcal{O}(\epsilon^2) \\ &= 1 + \epsilon n^\mu V(x)\Big(-\frac{\partial}{\partial x^\mu}V^\dagger(x)\Big) + \mathcal{O}(\epsilon^2). \end{aligned} \tag{15.26}$$

Then the terms proportional to ϵn^μ in (15.25) give the transformation

$$A^i_\mu(x)\frac{\sigma^i}{2} \to V(x)\Big(A^i_\mu(x)\frac{\sigma^i}{2} + \frac{i}{g}\partial_\mu\Big)V^\dagger(x). \tag{15.27}$$

The derivative acts on $V^\dagger(x) = \exp(-i\alpha^i\sigma^i/2)$; it is not so easy to compute this derivative explicitly because the exponent does not necessarily commute with its derivative. For infinitesimal transformations we can expand $V(x)$ to first order in α. In this case we obtain

$$A^i_\mu\frac{\sigma^i}{2} \to A^i_\mu\frac{\sigma^i}{2} + \frac{1}{g}(\partial_\mu\alpha^i)\frac{\sigma^i}{2} + i\big[\alpha^i\frac{\sigma^i}{2}, A^j_\mu\frac{\sigma^j}{2}\big] + \cdots. \tag{15.28}$$

The last term in this transformation law is new, and arises from the noncommutativity of the local transformations. By combining this relation with the infinitesimal form of the fermion transformation,

$$\psi \to \Big(1 + i\alpha^i\frac{\sigma^i}{2}\Big)\psi + \cdots, \tag{15.29}$$

we can check the infinitesimal transformation of the covariant derivative:

$$\begin{aligned} D_\mu\psi \to \Big(\partial_\mu - igA^i_\mu\frac{\sigma^i}{2} - i(\partial_\mu\alpha^i)\frac{\sigma^i}{2} + g\big[\alpha^i\frac{\sigma^i}{2}, A^j_\mu\frac{\sigma^j}{2}\big]\Big)\Big(1 + i\alpha^k\frac{\sigma^k}{2}\Big)\psi \\ = \Big(1 + i\alpha^i\frac{\sigma^i}{2}\Big)D_\mu\psi, \end{aligned} \tag{15.30}$$

up to terms of order α^2. It is not difficult to check using (15.27) and (15.21) that, even for finite transformations, the covariant derivative has the same transformation law as the field on which it acts.

Using the covariant derivative, we can build the most general gauge-invariant Lagrangians involving ψ. But to write a complete Lagrangian, we must also find gauge-invariant terms that depend only on A^i_μ. To do this, we construct the analogue of the electromagnetic field tensor. We will use the second method of the previous section, working from the commutator of covariant derivatives. The transformation law of the covariant derivative implies that

$$[D_\mu, D_\nu]\psi(x) \to V(x)[D_\mu, D_\nu]\psi(x). \tag{15.31}$$

At the same time, by writing out the commutator using formula (15.24), we can show, as in the Abelian case, that $[D_\mu, D_\nu]$ is not a differential operator but merely a multiplicative factor (now a matrix) acting on ψ. This time, however, there is a new feature: The last term in the expansion of the commutator no longer vanishes. Instead, we find

$$[D_\mu, D_\nu] = -igF^i_{\mu\nu}\frac{\sigma^i}{2}, \tag{15.32}$$

with

$$F^i_{\mu\nu}\frac{\sigma^i}{2} = \partial_\mu A^i_\nu\frac{\sigma^i}{2} - \partial_\nu A^i_\mu\frac{\sigma^i}{2} - ig\big[A^i_\mu\frac{\sigma^i}{2}, A^j_\nu\frac{\sigma^j}{2}\big]. \tag{15.33}$$

We can simplify this relation by applying the standard commutation relations of Pauli matrices:

$$\left[\frac{\sigma^i}{2}, \frac{\sigma^j}{2}\right] = i\epsilon^{ijk}\frac{\sigma^k}{2}. \tag{15.34}$$

Then

$$F^i_{\mu\nu} = \partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\epsilon^{ijk} A^j_\mu A^k_\nu. \tag{15.35}$$

The transformation law for the field strength follows from Eqs. (15.21) and (15.31):

$$F^i_{\mu\nu}\frac{\sigma^i}{2} \to V(x) F^j_{\mu\nu}\frac{\sigma^j}{2} V^\dagger(x). \tag{15.36}$$

The infinitesimal form is

$$F^i_{\mu\nu}\frac{\sigma^i}{2} \to F^i_{\mu\nu}\frac{\sigma^i}{2} + \left[i\alpha^i\frac{\sigma^i}{2}, F^j_{\mu\nu}\frac{\sigma^j}{2}\right]. \tag{15.37}$$

Notice that the field strength is no longer a gauge-invariant quantity. It cannot be, since there are now three field strengths, each associated with a given direction of rotation in the abstract space. However, it is easy to form gauge-invariant combinations of the field strengths. For example,

$$\mathcal{L} = -\frac{1}{2}\operatorname{tr}\left[\left(F^i_{\mu\nu}\frac{\sigma^i}{2}\right)^2\right] = -\frac{1}{4}\left(F^i_{\mu\nu}\right)^2 \tag{15.38}$$

is a gauge-invariant kinetic energy term for A^i_μ. Notice that, in contrast to the case of electrodynamics, this Lagrangian contains cubic and quartic terms in A^i_μ. Thus, this Lagrangian describes a nontrivial, interacting field theory, called *Yang-Mills theory.* This is the simplest example of a non-Abelian gauge theory.

To construct a theory of Yang-Mills vector fields interacting with fermions, we simply add the gauge-field Lagrangian (15.38) to the familiar Dirac Lagrangian, with the ordinary derivative of ψ replaced by the covariant derivative. The result looks almost identical to the QED Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\not{D})\psi - \frac{1}{4}(F^i_{\mu\nu})^2 - m\bar{\psi}\psi. \tag{15.39}$$

This is the famous Yang-Mills Lagrangian. Like that of QED, it depends on two parameters: the scale factor g (which is analogous to the electron charge) and the fermion mass m. By varying this Lagrangian, we find the classical equations of motion of the gauge theory. These are the Dirac equation for the fermion field and the equation

$$\partial^\mu F^i_{\mu\nu} + g\epsilon^{ijk} A^{j\mu} F^k_{\mu\nu} = -g\bar{\psi}\gamma_\nu\frac{\sigma^i}{2}\psi \tag{15.40}$$

for the vector field.

Everything that we have done for the $SU(2)$ symmetry transformation (15.20) generalizes easily to any other continuous group of symmetries. The

full range of possible symmetry groups is enumerated and classified in Section 15.4. For any such group, however, the general expressions for elements of the Lagrangian are quite similar. Consider any continuous group of transformations, represented by a set of $n \times n$ unitary matrices V. Then the basic fields $\psi(x)$ will form an n-plet, and transform according to

$$\psi(x) \to V(x)\psi(x), \tag{15.41}$$

where the x dependence of V makes the transformation local. In infinitesimal form, $V(x)$ can be expanded in terms of a set of basic generators of the symmetry group, which can be represented as Hermitian matrices t^a:

$$V(x) = 1 + i\alpha^a(x)t^a + \mathcal{O}(\alpha^2). \tag{15.42}$$

Now one can carry through the whole analysis from Eq. (15.22) to Eq. (15.33) for a general local symmetry group simply by replacing

$$\frac{\sigma^i}{2} \to t^a \tag{15.43}$$

at each step of the analysis.

To generalize the explicit expression (15.35) for the field tensor, we need to know the commutation relations of the matrices t^a. It is conventional to write these in the standard form

$$[t^a, t^b] = if^{abc}t^c, \tag{15.44}$$

where f^{abc} is a set of numbers called *structure constants*. This object replaces ϵ^{ijk} in Eq. (15.34). It is conventional to choose a basis for the matrices t^a such that f^{abc} is completely antisymmetric; we will prove that this is always possible in Section 15.4.

We can now recapitulate all of our results as follows. The covariant derivative associated with the general transformation (15.41) is

$$D_\mu = \partial_\mu - igA^a_\mu t^a; \tag{15.45}$$

it contains one vector field for each independent generator of the local symmetry. The infinitesimal tranformation laws for ψ and A^a_μ are

$$\begin{aligned} \psi &\to (1 + i\alpha^a t^a)\psi; \\ A^a_\mu &\to A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc}A^b_\mu \alpha^c. \end{aligned} \tag{15.46}$$

The finite transformation of A^a_μ has exactly the form of (15.27):

$$A^a_\mu(x)t^a \to V(x)\Big(A^a_\mu(x)t^a + \frac{i}{g}\partial_\mu\Big)V^\dagger(x). \tag{15.47}$$

These transformation laws imply that the covariant derivative of ψ has the same transformation law as ψ itself. The field tensor is defined by

$$[D_\mu, D_\nu] = -igF^a_{\mu\nu}t^a, \tag{15.48}$$

or more explicitly,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu. \tag{15.49}$$

This quantity has the infinitesimal transformation

$$F^a_{\mu\nu} \to F^a_{\mu\nu} - f^{abc}\alpha^b F^c_{\mu\nu}. \tag{15.50}$$

From Eqs. (15.46) and (15.50), one can show that any globally symmetric function of ψ, $F^a_{\mu\nu}$, and their covariant derivatives is also locally symmetric, and is therefore a candidate for a term in a gauge-invariant Lagrangian. However, there are very few permissible terms up to dimension 4. The most general gauge-invariant Lagrangian that is renormalizable and conserves P and T is again given by Eq. (15.39). The corresponding classical equation of motion is

$$\partial^\mu F^a_{\mu\nu} + g f^{abc} A^{b\mu} F^c_{\mu\nu} = -g j^a_\nu, \tag{15.51}$$

where

$$j^a_\nu = \bar{\psi}\gamma_\nu t^a \psi \tag{15.52}$$

is the global symmetry current of the fermion field.

Notice that the nonlinear terms in the Yang-Mills Lagrangian (15.39) appear in the covariant derivative, where they are proportional to t^a, and in the field tensor, where they are proportional to f^{abc}. Thus the form of the interactions in a non-Abelian gauge theory is dictated by the local symmetry. The nonlinear interactions of the vector field with itself are proportional to the commutators of symmetry generators and thus explicitly require the non-Abelian nature of the symmetry group.

15.3 The Gauge-Invariant Wilson Loop

In both of the previous sections we made use of the *comparator*, $U(y,x)$, which converts the fermion gauge transformation law at point x to that at point y. So far, in writing expressions for this object, it has sufficed to assume that x and y are infinitesimally separated. However, it is worthwhile to think further about the comparator in the case where x and y are far apart. This discussion will give us further insights into the geometry of gauge invariance, and will reveal some additional useful functions of the gauge field which we will put to work in Chapter 19.

We first return to the Abelian theory and expand upon our discussion of $U(y,x)$ in that context. In Eq. (15.10) we constructed a product of comparators on a path that wound around a small square. We showed that this product $\mathbf{U}(x)$ is not trivial, even though we eventually return to the starting point; rather, we found that $\mathbf{U}(x)$ differs from 1 by a term proportional to the electromagnetic field strength and to the area of the square. This is a particular case of a general conclusion: The comparator between two points x and y at finite separation depends on the path taken from x to y.

To explain this statement, it is useful to reverse some of the logic of Section 15.1. We begin from the connection A_μ, which we assume to have the transformation law (15.7), and construct $U(z,y)$ as a function of A_μ that transforms according to (15.3). It is not difficult to verify that the expression

$$U_P(z,y) = \exp\Big[-ie\int_P dx^\mu A_\mu(x)\Big] \tag{15.53}$$

meets this criterion if the integral is taken along any path P that runs from y to z. This object $U_P(z,y)$ is called the *Wilson line.*‡ Expression (15.53) gives an explicit realization of the abstract comparator $U(z,y)$ for points at finite separation.

A crucial property of the Wilson line is that it depends on the path P. If P is a closed path that returns to y, we obtain the *Wilson loop*,

$$U_P(y,y) = \exp\Big[-ie\oint_P dx^\mu A_\mu(x)\Big]. \tag{15.54}$$

This quantity is a nontrivial function of A_μ that is, by construction, locally gauge invariant. In fact, all gauge-invariant functions of A_μ can be thought of as combinations of Wilson loops for various choices of the path P. To motivate this claim, we use Stokes's theorem to rewrite the Wilson loop as

$$U_P(y,y) = \exp\Big[-i\frac{e}{2}\int_\Sigma d\sigma^{\mu\nu}F_{\mu\nu}\Big], \tag{15.55}$$

where Σ is a surface that spans the closed loop P, $d\sigma^{\mu\nu}$ is an area element on this surface, and $F_{\mu\nu}$ is the field tensor (15.13). This relation between the Wilson loop and the field strength is illustrated in Fig. 15.2. Since the Wilson loop is gauge invariant, this argument gives one more way to visualize the gauge invariance of the field strength. Conversely, since (almost) all gauge-invariant functions of A_μ can be built up from $F_{\mu\nu}$, this expression gives weight to the statement that $U_P(y,y)$ is the most general gauge invariant.

Both the Wilson line and the Wilson loop can be generalized to the non-Abelian case. Here, however, additional subtleties arise when we consider exponentials of noncommuting matrices. Let us first construct the Wilson line, which now transforms according to Eq. (15.22). It is not correct to make a straightforward rewriting of (15.53) with the integral of $A^a_\mu t^a$ in the exponent, since these matrices do not necessarily commute at different points. Instead, we must order these matrices in a particular way. We will now give the correct ordering prescription and then prove its transformation law.

Let s be a parameter of the path P, running from 0 at $x=y$ to 1 at $x=z$. Then define the Wilson line as the power-series expansion of the exponential, with the matrices in each term ordered so that higher values of s stand to the

‡This path-dependent phase was used long before Wilson's work, in Schwinger's early papers on QED, and in Y. Aharonov and D. Böhm, *Phys. Rev.* **115**, 485 (1959).

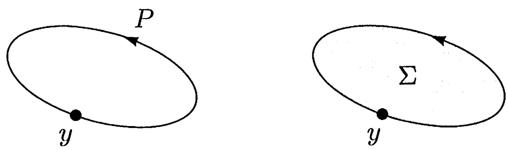


Figure 15.2. The Wilson loop integral is taken around an arbitrary loop. It can also be expressed as a flux integral of the field strength over a surface spanning the loop.

left. This prescription is called *path-ordering* and is denoted by the symbol $P\{\}$. Thus the Wilson line is written

$$U_P(z,y) = P\left\{\exp\left[ig\int_0^1 ds\frac{dx^\mu}{ds}A^a_\mu(x(s))t^a\right]\right\}. \tag{15.56}$$

This expression is similar to the time-ordered exponential that we wrote for the interaction-picture propagator in Eq. (4.23). Pursuing this analogy, one can show that this expression for U_P is the solution of a differential equation similar to (4.24):

$$\frac{d}{ds}U_P(x(s),y) = \left(ig\frac{dx^\mu}{ds}A^a_\mu(x(s))t^a\right)U_P(x(s),y). \tag{15.57}$$

(Here we consider U_P to be a continuous function of the parameter s, rather than fixing $s=1$ at the endpoint.)

To show that expression (15.56) is the correct generalization of the Wilson line, we must show that it satisfies the correct gauge transformation law (15.22). This follows from the differential equation (15.57), which can be rewritten as

$$\frac{dx^\mu}{ds}\,D_\mu U_P(x,y) = 0. \tag{15.58}$$

Now let A^V represent the gauge transform of a field configuration A, and use these arguments to denote explicitly the dependence of gauge functions on the gauge field. We would like to show that

$$U_P(z,y,A^V) = V(z)U_P(z,y,A)V^\dagger(y), \tag{15.59}$$

which is equivalent to (15.22). In (15.30) we proved, in its infinitesimal version, the relation

$$D_\mu(A^V)\,V(x) = V(x)\,D_\mu(A). \tag{15.60}$$

This relation implies that the right-hand side of (15.59) satisfies (15.58) for the gauge field A^V if $U_P(z,y,A)$ satisfies this equation for the gauge field A. But the solution of a first-order differential equation with a fixed boundary condition is unique. Thus, if $U_P(z,y)$ is defined to be the solution of (15.57) or (15.58), it indeed has the transformation law (15.59).

The Wilson line associated with a closed path returning to y transforms only with the gauge parameter at y; however, it is not a gauge invariant:

$$U_P(y,y) \to V(y)U_P(y,y)V^\dagger(y). \tag{15.61}$$

To understand this transformation better, one can work out the expression for $U_P(x,x)$, where the path is the small square in the $(1,2)$ plane shown in Fig. 15.1. In addition to the terms in Eq. (15.11), there are additional corrections of order ϵ^2 coming from products of $(A^a_\mu t^a)$ factors from pairs of sides, which sum up to a commutator of these factors. One finds

$$U_P(x,x) = 1 + ig\epsilon^2 F^a_{12}(x)t^a + \mathcal{O}(\epsilon^3), \tag{15.62}$$

where $F^a_{\mu\nu}$ is given by the full expression in (15.49). If we then expand the transformation law (15.61) in powers of ϵ, the term of order ϵ^2 is the transformation law of $F^a_{\mu\nu}$ given in Eq. (15.36).

To convert the Wilson line for a closed path into a true gauge invariant, take the trace. By cyclic invariance, (15.61) implies

$$\operatorname{tr} U_P(x,x) \to \operatorname{tr} U_P(x,x). \tag{15.63}$$

Thus for a non-Abelian gauge theory, we define the Wilson loop to be the trace of the Wilson line around a closed path.

Let us evaluate $\operatorname{tr} U_P(x,x)$ more explicitly for the case of an $SU(2)$ gauge group. If $U(\epsilon)$ is any 2×2 unitary matrix that tends to 1 as $\epsilon \to 0$, we can expand it in ϵ as follows:

$$\begin{aligned} U(\epsilon) &= \exp\left[i(\epsilon\beta^i + \epsilon^2\gamma^i + \cdots)\frac{\sigma^i}{2}\right] \\ &= 1 + i(\epsilon\beta^i + \epsilon^2\gamma^i + \cdots)\frac{\sigma^i}{2} - \frac{1}{2}(\epsilon\beta^i \cdot \epsilon\beta^j + \cdots)\frac{\sigma^i}{2}\frac{\sigma^j}{2} + \cdots . \end{aligned} \tag{15.64}$$

Then, since the Pauli matrices are traceless and satisfy $\operatorname{tr}[\sigma^i\sigma^j] = 2\delta^{ij}$,

$$\operatorname{tr} U(\epsilon) = 2 - \frac{1}{4}\epsilon^2(\beta^i)^2 + \mathcal{O}(\epsilon^3). \tag{15.65}$$

Applying this formula to Eq. (15.62), we find

$$\operatorname{tr} U_P(x,x) = 2 - \frac{1}{4}g^2\epsilon^4(F^i_{12})^2 + \mathcal{O}(\epsilon^5). \tag{15.66}$$

Thus the gauge invariance of $(F^i_{\mu\nu})^2$ can be derived from a geometrical argument, just as in the Abelian case. Using the notation that will be introduced in the next section, one can show that the same argument goes through for any gauge group.

15.4 Basic Facts about Lie Algebras

At the end of Section 15.2 we saw that the class of non-Abelian gauge theories is very large. To work with these theories most efficiently, it is worthwhile to pause and consider the general properties of the continuous groups on which they are based. In this section we will enumerate all the possible groups that can be used to construct non-Abelian gauge theories. We will then compute some numerical factors, built out of group transformation matrices, that are needed in performing explicit calculations in quantized gauge theories.*

To a mathematician, a group is made up of abstract entities that obey certain algebraic rules. In quantum mechanics, however, we are interested specifically in groups of unitary operators that act on the vector space of quantum states. We focus our attention on continuously generated groups, that is, groups that contain elements arbitrarily close to the identity, such that the general element can be reached by the repeated action of these infinitesimal elements. Then any infinitesimal group element g can be written

$$g(\alpha) = 1 + i\alpha^a T^a + \mathcal{O}(\alpha^2). \tag{15.67}$$

The coefficients of the infinitesimal group parameters α^a are Hermitian operators T^a, called the *generators* of the symmetry group. A continuous group with this structure is called a *Lie group.*

The set of generators T^a must span the space of infinitesimal group transformations, so the commutator of generators T^a must be a linear combination of generators. Thus the commutation relations of the operators T^a can be written

$$[T^a, T^b] = if^{abc}T^c; \tag{15.68}$$

the numbers f^{abc} are called *structure constants.* The vector space spanned by the generators, with the additional operation of commutation, is called a *Lie Algebra.*

The commutation relations (15.68) and the identity

$$[T^a, [T^b, T^c]] + [T^b, [T^c, T^a]] + [T^c, [T^a, T^b]] = 0 \tag{15.69}$$

imply that the structure constants obey

$$f^{ade}f^{bcd} + f^{bde}f^{cad} + f^{cde}f^{abd} = 0, \tag{15.70}$$

called the *Jacobi identity.* From the mathematician's viewpoint (considering the generators to be abstract entities rather than Hermitian operators), the

*In this section we will state, without proof, some general results from the theory of continuous groups. There are many excellent books that review these mathematical results systematically. Among these, we recommend especially Cahn (1984), for a brief but incisive discussion, and S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic Press, 1978), which gives an elegant and rigorous account. R. Slansky, *Phys. Repts.* **79**, 1 (1981), has compiled an especially useful set of tables of group-theoretic identities relevant to the construction of non-Abelian gauge theories.

Jacobi identity is an axiom that must be satisfied in order for a given set of commutation rules to define a Lie algebra.

The commutation relations of the Lie algebra completely determine the group multiplication law of an associated Lie group sufficiently close to the identity. For large enough transformations, additional global questions come into play; to give a familiar example, $SU(2)$ and $O(3)$ have the same commutation relations but different global structure. However, the Lagrangian of a non-Abelian gauge theory depends only on the Lie algebra of the local symmetry group, so we will ignore these global questions from here on.

Classification of Lie Algebras

For the application to gauge theories, the local symmetry is normally a unitary transformation of a set of fields. Thus we are primarily interested in Lie algebras that have finite-dimensional Hermitian representations, leading to finite-dimensional unitary representations of the corresponding Lie group. We will also assume that the number of generators is finite. Such Lie algebras are called *compact*, because these conditions imply that the Lie group is a finite-dimensional compact manifold.

If one of the generators T^a commutes with all of the others, it generates an independent continuous Abelian group. Such a group, which has the structure of the group of phase rotations

$$\psi \to e^{i\alpha}\psi, \tag{15.71}$$

we call $U(1)$. If the algebra contains no such commuting elements, so that the group contains no $U(1)$ factors, then we call the algebra *semi-simple*. If, in addition, the Lie algebra cannot be divided into two mutually commuting sets of generators, the algebra is *simple*. A general Lie algebra is the direct sum of non-Abelian simple components and additional Abelian generators.

Surprisingly, the basic conditions that a Lie algebra be compact and simple turn out to be extremely restrictive. In one of the triumphs of nineteenth-century mathematics, Killing and Cartan classified all possible compact simple Lie algebras. Almost all of these algebras belong to one of three infinite families, with only five exceptions. The three infinite families are the algebras corresponding to the so-called *classical groups*, whose structures are conveniently defined in terms of particular matrix representations. The definitions of the three families of classical groups are as follows:

1. *Unitary transformations of N-dimensional vectors.* Let ξ and η be complex N-vectors. A general linear transformation then has the form

$$\eta_a \to U_{ab}\eta_b, \qquad \xi_a \to U_{ab}\xi_b. \tag{15.72}$$

We say that this transformation is *unitary* if it preserves the inner product $\eta_a^*\xi_a$. The pure phase transformations

$$\xi_a \to e^{i\alpha}\xi_a \tag{15.73}$$

form a $U(1)$ subgroup which commutes with all other unitary transformations; we remove this subgroup to form a simple Lie group, called $SU(N)$; it consists of all $N \times N$ unitary transformations satisfying $\det(U) = 1$. The generators of $SU(N)$ are represented by $N \times N$ Hermitian matrices t^a, subject to the condition that they be orthogonal to the generator of (15.73):

$$\mathrm{tr}[t^a] = 0. \tag{15.74}$$

There are $N^2 - 1$ independent matrices satisfying these conditions.

2. *Orthogonal transformations of N-dimensional vectors.* This is the subgroup of unitary $N \times N$ transformations that preserves the symmetric inner product

$$\eta_a E_{ab} \xi_b, \qquad \text{with } E_{ab} = \delta_{ab}. \tag{15.75}$$

This is the usual vector product, and so this group is the rotation group in N dimensions, $SO(N)$. (Adding the reflection gives the group $O(N)$.) There is an independent rotation corresponding to each plane in N dimensions, so $SO(N)$ has $N(N-1)/2$ generators.

3. *Symplectic transformations of N-dimensional vectors.* This is the subgroup of unitary $N \times N$ transformations, for N even, that preserves the antisymmetric inner product

$$\eta_a E_{ab} \xi_b, \qquad \text{with } E_{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{15.76}$$

where the elements of the matrix are $N/2 \times N/2$ blocks. This group is called $Sp(N)$; it has $N(N+1)/2$ generators.

Beyond these three families, there are five more *exceptional* Lie algebras, denoted in Cartan's classification system as G_2, F_4, E_6, E_7, and E_8. Of these, E_6 and E_8 have been applied as local symmetry groups in interesting unified models of the fundamental interactions. However, we will not consider these exceptional groups further in this book. In fact, most of our examples will involve only $SU(N)$ groups.

Representations

Once we have specified the local symmetry group, the fields that appear in the Lagrangian most naturally transform according to a finite-dimensional unitary representation of this group. Thus we might next ask how to systematically find all such representations of any given Lie group. Recall that for the group $SU(2)$, the representations can be constructed directly from the commutation relations, using the raising and lowering operators J_+ and J_-. This construction can be generalized to find the finite-dimensional representations of any compact Lie algebra. In this book, however, we will work with relatively simple representations whose structure we can work out by less formal methods.

Before discussing representations of Lie algebras, we should review some general aspects of group representations. Given a symmetry group G, a finite-dimensional unitary representation of the group's Lie algebra is a set of $d \times d$ Hermitian matrices t^a that satisfy the commutation relations (15.68). The size d is the *dimension* of the representation. An arbitrary representation can generally be decomposed by finding a basis in which all representation matrices are simultaneously block-diagonal. Through this change of basis, we can write the representation as the direct sum of *irreducible* representations. We denote the representation matrices in the irreducible representation r by t^a_r.

It is standard practice to adopt a normalization convention for the matrices t^a_r, based on traces of their products. If the Lie algebra is semi-simple, the matrices t^a_r themselves are traceless. Consider, however, the trace of the product of two generator matrices:

$$\operatorname{tr}[t^a_r t^b_r] \equiv D^{ab}. \tag{15.77}$$

As long as the generator matrices are Hermitian, the matrix D^{ab} is positive definite. Let us choose a basis for the generators T^a so that this matrix is proportional to the identity. It can be shown that, once this is done for one irreducible representation, it is true for all irreducible representations. Thus, in this basis,

$$\operatorname{tr}[t^a_r t^b_r] = C(r)\delta^{ab}, \tag{15.78}$$

where $C(r)$ is a constant for each representation r. Equation (15.78) and the commutation relations (15.68) yield the following representation of the structure constants:

$$f^{abc} = -\frac{i}{C(r)} \operatorname{tr}\left\{[t^a_r, t^b_r]t^c_r\right\}. \tag{15.79}$$

This equation implies that f^{abc} is totally antisymmetric.

For each irreducible representation r of G, there is an associated *conjugate* representation $\bar{r}$. The representation r yields the infinitesimal transformation

$$\phi \to (1 + i\alpha^a t^a_r)\phi. \tag{15.80}$$

The complex conjugate of this transformation,

$$\phi^* \to (1 - i\alpha^a (t^a_r)^*)\phi^*, \tag{15.81}$$

must also be the infinitesimal element of a representation of G. Thus the conjugate representation to r has representation matrices

$$t^a_{\bar{r}} = -(t^a_r)^* = -(t^a_r)^T. \tag{15.82}$$

Since $\phi^*\phi$ is invariant to unitary transformations, it is possible to combine fields transforming in the representations r and $\bar{r}$ to form a group invariant.

It is possible that the representation $\bar{r}$ may be equivalent to r, if there is a unitary transformation U such that $t^a_{\bar{r}} = U t^a_r U^\dagger$. If so, the representation r is *real*. In this case, there is a matrix G_{ab} such that, if η and ξ belong to the representation r, the combination $G_{ab}\eta_a\xi_b$ is an invariant. It is sometimes

useful to distinguish the case in which G_{ab} is symmetric from that in which G_{ab} is antisymmetric. In the former case the representation is *strictly real*; in the latter case it is *pseudoreal.* Both cases occur already in $SU(2)$: The invariant combination of two vectors is $v_a w_a$, so the vector is a real representation; the invariant combination of two spinors is $\epsilon^{\alpha\beta}\eta_\alpha\xi_\beta$, so the spinor is a pseudoreal representation.

With this language we can discuss the simplest representations of the classical groups. In $SU(N)$, the basic irreducible representation (often called the *fundamental* representation) is the N-dimensional complex vector. For $N > 2$ this representation is complex, so that there is a second, inequivalent, representation $\overline{N}$. (In $SU(2)$ this representation is the pseudoreal spinor representation.) In $SO(N)$, the basic N-dimensional vector is a (strictly) real representation. In $Sp(N)$, the N-dimensional vector is a pseudoreal representation.

Another irreducible representation, present for any simple Lie algebra, is the one to which the generators of the algebra belong. This representation is called the *adjoint representation* and denoted by $r = G$. The representation matrices are given by the structure constants:

$$(t_G^b)_{ac} = if^{abc}. \tag{15.83}$$

With this definition, the statement that t_G^a satisfies the Lie algebra

$$\left([t_G^b, t_G^c]\right)_{ae} = if^{bcd}(t_G^d)_{ae} \tag{15.84}$$

is just a rewriting of the Jacobi identity (15.70). Since the structure constants are real and antisymmetric, $t_G^a = -(t_G^a)^*$; thus the adjoint representation is always a real representation. From the descriptions of the Lie groups given above, the dimension of the adjoint representation $d(G)$ is given, for the classical groups, by

$$d(G) = \begin{cases} N^2 - 1 & \text{for } SU(N), \\ N(N-1)/2 & \text{for } SO(N), \\ N(N+1)/2 & \text{for } Sp(N). \end{cases} \tag{15.85}$$

The identification of f^{abc} as a representation matrix allows us to gain further insight into some of the quantities introduced in Section 15.2. The covariant derivative acting on a field in the adjoint representation is

$$\begin{aligned}(D_\mu\phi)_a &= \partial_\mu\phi_a - igA_\mu^b(t_G^b)_{ac}\phi_c \\ &= \partial_\mu\phi_a + gf^{abc}A_\mu^b\phi_c.\end{aligned} \tag{15.86}$$

Thus we can recognize the infinitesimal form of the gauge transformation of the vector field in (15.46) as the motion

$$A_\mu^a \to A_\mu^a + \frac{1}{g}(D_\mu\alpha)^a. \tag{15.87}$$

The gauge field equation of motion (15.51) can be rewritten as

$$(D^\mu F_{\mu\nu})^a = -gj^a_\nu. \tag{15.88}$$

In both of these expressions, the arbitrary-looking terms involving f^{abc} arise naturally as part of a covariant derivative. An additional identity follows from considering the antisymmetric double commutator of covariant derivatives,

$$\epsilon^{\mu\nu\lambda\sigma}[D_\nu, [D_\lambda, D_\sigma]].$$

This quantity vanishes by its total antisymmetry, in the same way as (15.69). This result can be reduced to the identity

$$\epsilon^{\mu\nu\lambda\sigma}(D_\nu F_{\lambda\sigma})^a = 0. \tag{15.89}$$

This equation, called the *Bianchi identity* of a non-Abelian gauge theory, is the analogue of the homogeneous Maxwell equations in electrodynamics.

The Casimir Operator

In $SU(2)$, we characterize representations by the eigenvalue of the total spin J^2. In fact, for any simple Lie algebra, the operator

$$T^2 = T^a T^a \tag{15.90}$$

(with the repeated index summed, as always) commutes with all group generators:

$$\begin{aligned}[][T^b, T^a T^a] &= (if^{bac}T^c)T^a + T^a(if^{bac}T^c) \\ &= if^{bac}\{T^c, T^a\},\end{aligned} \tag{15.91}$$

which vanishes by the antisymmetry of f^{abc}. In other words, T^2 is an invariant of the algebra; this implies that T^2 takes a constant value on each irreducible representation. Thus, the matrix representation of T^2 is proportional to the unit matrix:

$$t^a_r t^a_r = C_2(r)\cdot \mathbf{1}, \tag{15.92}$$

where $\mathbf{1}$ is the $d(r)\times d(r)$ unit matrix and $C_2(r)$ is a constant, called the *quadratic Casimir operator*, for each representation. For the adjoint representation, Eq. (15.92) is more conveniently written as

$$f^{acd}f^{bcd} = C_2(G)\delta^{ab}. \tag{15.93}$$

Casimir operators appear very often in computations in non-Abelian gauge theories. Furthermore, the related invariant $C(r)$ given by (15.78) is simply related to the Casimir operator: If we contract (15.78) with δ^{ab} and evaluate the left-hand side using (15.92), we find

$$d(r)C_2(r) = d(G)C(r). \tag{15.94}$$

Thus it will be useful for us to compute $C_2(r)$ for the simplest $SU(N)$ representations.

For $SU(2)$, the fundamental two-dimensional representation is the spinor representation, which is given in terms of Pauli matrices by

$$t_2^a = \frac{\sigma^a}{2}. \tag{15.95}$$

These satisfy $\text{tr}[t_2^a t_2^b] = \frac{1}{2}\delta^{ab}$. We will choose the generators of $SU(N)$ so that three of these are the generators (15.95), acting on the first two components of the N-vector ξ. Then, for any matrices of the fundamental representation,

$$\text{tr}[t_N^a t_N^b] = \tfrac{1}{2}\delta^{ab}. \tag{15.96}$$

This convention fixes the values of $C(r)$ and $C_2(r)$ for all of the irreducible representations of $SU(N)$. For the fundamental representations N and $\overline{N}$, $C(N)$ is given directly by (15.96), and $C_2(N)$ follows from (15.94). We find

$$C(N) = \frac{1}{2}, \qquad C_2(N) = \frac{N^2-1}{2N}. \tag{15.97}$$

To compute the Casimir operator for the adjoint representation, we build up this representation from the product of the N and $\overline{N}$. Let us first discuss the product of irreducible representations more generally. The direct product of two representations r_1, r_2 is a representation of dimension $d(r_1)\cdot d(r_2)$. An object that transforms according to this representation can be written as a tensor Ξ_{pq}, in which the first index transforms according to r_1, the second according to r_2. In general, such a product can be decomposed into a direct sum of irreducible representations; symbolically, we write

$$r_1 \times r_2 = \sum r_i. \tag{15.98}$$

The representation matrices in the representation $r_1 \times r_2$ are

$$t^a_{r_1\times r_2} = t^a_{r_1} \otimes 1 + 1 \otimes t^a_{r_2}, \tag{15.99}$$

where the first matrix of each product acts on the first index of Ξ_{pq} and the second matrix acts on the second index.

The Casimir operator in the product representation is

$$(t^a_{r_1\times r_2})^2 = (t^a_{r_1})^2 \otimes 1 + 2t^a_{r_1} \otimes t^a_{r_2} + 1 \otimes (t^a_{r_2})^2.$$

Take the trace; since the matrices t^a_r are traceless, the trace of the second term on the right is zero. Then

$$\text{tr}(t^a_{r_1\times r_2})^2 = \big(C_2(r_1) + C_2(r_2)\big)d(r_1)d(r_2). \tag{15.100}$$

On the other hand, the decomposition (15.98) implies

$$\text{tr}(t^a_{r_1\times r_2})^2 = \sum C_2(r_i)d(r_i). \tag{15.101}$$

Equating (15.100) and (15.101), we find a useful identity for $C_2(r)$.

Now apply this identity to the product of the N and $\overline{N}$ representations of $SU(N)$. In this case, the tensor Ξ_{pq} can contain a term proportional to the invariant δ_{pq}. The remaining (N^2-1) independent components of Ξ_{pq}

transform as a general traceless $N \times N$ tensor; the matrices that effect these transformations make up the adjoint representation of $SU(N)$. In this case Eq. (15.98) becomes explicitly

$$N \times \overline{N} = 1 + (N^2 - 1). \tag{15.102}$$

For this decomposition, Eqs. (15.100) and (15.101) imply the identity

$$\left(2 \cdot \frac{N^2 - 1}{2N}\right) N^2 = 0 + C_2(G) \cdot (N^2 - 1). \tag{15.103}$$

Thus, for $SU(N)$,

$$C_2(G) = C(G) = N. \tag{15.104}$$

Some additional examples of the computation of quadratic Casimir operators are given in Problem 15.5. However, the examples we have discussed in this section, combined with the basic group-theoretic concepts that we have reviewed, already provide enough material to carry out the most important computations of physical interest in non-Abelian gauge theories.

Problems

15.1 Brute-force computations in $SU(3)$. The standard basis for the fundamental representation of $SU(3)$ is

$$t^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad t^2 = \frac{1}{2}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad t^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$t^4 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad t^5 = \frac{1}{2}\begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$

$$t^6 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad t^7 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad t^8 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

(a) Explain why there are exactly eight matrices in the basis.

(b) Evaluate all the commutators of these matrices, to determine the structure constants of $SU(3)$. Show that, with the normalizations used here, f^{abc} is totally antisymmetric. (This exercise is tedious; you may wish to check only a representative sample of the commutators.)

(c) Check the orthogonality condition (15.78), and evaluate the constant $C(r)$ for this representation.

(d) Compute the quadratic Casimir operator $C_2(r)$ directly from its definition (15.92), and verify the relation (15.94) between $C_2(r)$ and $C(r)$.

15.2 Write down the basis matrices of the adjoint representation of $SU(2)$. Compute $C(G)$ and $C_2(G)$ directly from their definitions (15.78) and (15.92).

15.3 Coulomb potential.

(a) Using functional integration, compute the expectation value of the Wilson loop in pure quantum electrodynamics without fermions. Show that

$$\langle U_P(z,z)\rangle = \exp\left[-e^2 \oint_P dx^\mu \oint_P dy^\nu g_{\mu\nu} \frac{1}{8\pi^2(x-y)^2}\right],$$

with x and y integrated around the closed curve P.

(b) Consider the Wilson loop of a rectangular path of (spacelike) width R and (timelike) length T, $T \gg R$. Compute the expectation value of the Wilson loop in this limit and compare to the general expression for time evolution,

$$\langle U_P\rangle = \exp[-iE(R)T],$$

where $E(R)$ is the energy of the electromagnetic sources corresponding to the Wilson loop. Show that the potential energy of these sources is just the Coulomb potential, $V(R) = -e^2/4\pi R$.

(c) Assuming that the propagator of the non-Abelian gauge field is given by the Feynman gauge expression

$$\left\langle A^a_\mu(x) A^b_\mu(y)\right\rangle = \int \frac{d^4p}{(2\pi)^4} \frac{-ig_{\mu\nu}\delta^{ab}}{p^2} e^{-ip\cdot(x-y)},$$

compute the expectation value of a non-Abelian Wilson loop to order g^2. The result will depend on the representation r of the gauge group in which one chooses the matrices that appear in the exponential. Show that, to this order, the Coulomb potential of the non-Abelian gauge theory is $V(R) = -g^2 C_2(r)/4\pi R$.

15.4 Scalar propagator in a gauge theory. Consider the equation for the Green's function of the Klein-Gordon equation:

$$(\partial^2 + m^2) D_F(x,y) = -i\delta^{(4)}(x-y).$$

We can find an interesting representation for this Green's function by writing

$$D_F(x,y) = \int_0^\infty dT\, D(x,y,T),$$

where $D(x,y,T)$ satisfies the Schrödinger equation

$$\left[i\frac{\partial}{\partial T} - (\partial^2 + m^2)\right] D(x,y,T) = i\delta(T)\delta^{(4)}(x-y).$$

Now, represent $D(x,y,T)$ using the functional integral solution of the Schrödinger equation presented in Section 9.1.

(a) Using the explicit formula of the propagator of the Schrödinger equation, show that this integral formula gives the standard expression for the Feynman propagator.

(b) Using the method just described, show that the expression

$$D_F(x,y) = \int_0^\infty dT \int \mathcal{D}x \, \exp\left[i \int dt \, \frac{1}{2}\left(\left(\frac{dx^\mu}{dt}\right)^2 - m^2 \right) - ie \int dt \, \frac{dx^\mu}{dt} A_\mu(x) \right]$$

is a functional integral representation for the scalar field propagator in an arbitrary background electromagnetic field. Show, in particular, that the functional integral satisfies the relevant Schrödinger equation. Notice that this integral depends on A_μ through the Wilson line.

(c) Generalize this expression to a non-Abelian gauge theory. Show that the functional integral solves the relevant Schrödinger equation only if the group matrices in the exponential for the Wilson line are path-ordered.

15.5 Casimir operator computations. An alternative strategy for computing the quadratic Casimir operator is to compute $C(r)$ in the formula

$$\mathrm{tr}[t_r^a t_r^b] = C(r)\delta^{ab}$$

by choosing t^a and t^b to lie in an $SU(2)$ subgroup of the gauge group.

(a) Under an $SU(2)$ subgroup of a general group G, an irreducible representation r of G will decompose into a sum of representations of $SU(2)$:

$$r \to \sum j_i,$$

where the j_i are the spins of $SU(2)$ representations. Show that

$$3C(r) = \sum_i j_i(j_i+1)(2j_i+1).$$

(b) Under an $SU(2)$ subgroup of $SU(N)$, the fundamental representation N transforms as a 2-component spinor ($j = \frac{1}{2}$) and $(N-2)$ singlets. Use this relation to check the formula $C(N) = \frac{1}{2}$. Show that the adjoint representation of $SU(N)$ decomposes into one spin 1, $2(N-2)$ spin-$\frac{1}{2}$'s, plus singlets, and use this decomposition to check that $C(G) = N$.

(c) Symmetric and antisymmetric 2-index tensors form irreducible representations of $SU(N)$. Compute $C_2(r)$ for each of these representations. The direct sum of these representations is the product representation $N \times N$. Verify that your results for $C_2(r)$ satisfy the identity for product representations that follows from Eqs. (15.100) and (15.101).

Chapter 16

Quantization of Non-Abelian Gauge Theories

The previous chapter showed how to construct Lagrangians with non-Abelian gauge symmetry. However, this is only the first step in the process of relating the idea of non-Abelian gauge invariance to the real interactions of particle physics. We must next work out the rules for computing Feynman diagrams containing the non-Abelian gauge vector particles, then use these rules to compute scattering amplitudes and cross sections. This chapter will develop the technology needed for such calculations.

Alongside this technical discussion, we will study how the gauge symmetry affects the Feynman amplitudes. In any theory with a local symmetry, some degrees of freedom of the fields that appear in the Lagrangian are *unphysical*, in the sense that they can be adjusted arbitrarily by gauge transformations. In electrodynamics, the components of the field $A_\mu(k)$ proportional to k^μ lie along the symmetry directions. We saw in Section 9.4 that this fact has two important consequences. First, the propagator of the field A_μ is ambiguous; there are multiple expressions for the propagator, which follow equally well from the QED Lagrangian. Second, the vertices of electrodynamics are such that this ambiguity makes no difference in the calculation of cross sections. For example, Eq. (9.58) displays a continuous family of photon propagators, one for each value of the continuous parameter ξ; but we saw immediately that all dependence of S-matrix elements on ξ is eliminated by the Ward identity. Non-Abelian gauge theories contain similar ambiguities and cancellations, but, as we will see in this chapter, the structure of the cancellations is more intricate.

An additional goal of this chapter is to compute the Callan-Symanzik β function, and hence determine the behavior of the running coupling constant, for non-Abelian gauge theories. As discussed in Chapter 14, these theories are in fact *asymptotically free*: The coupling constant becomes weak at large momenta. This result indicates the applicability of non-Abelian gauge theory to model the strong interactions. We will be able to derive this result once we have determined the correct Feynman rules for non-Abelian gauge theories.

16.1 Interactions of Non-Abelian Gauge Bosons

Most of the Feynman rules for non-Abelian gauge theory can be read directly from the Yang-Mills Lagrangian, following the method of Section 9.2. However, when we quantized the electromagnetic field in Section 9.4, we saw that the functional integral over a gauge field must be defined carefully, and that the subtle aspects of this construction can introduce new ingredients into the quantum theory. In this section we will see how far we can go in the non-Abelian theory by ignoring these subtleties. In Section 16.2 we will carry out a more proper derivation of the Feynman rules, through a careful analysis of the functional integral.

Feynman Rules for Fermions and Gauge Bosons

The Yang-Mills Lagrangian, as derived in the previous chapter, is

$$\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2 + \bar{\psi}(i\not{D} - m)\psi, \tag{16.1}$$

where the index a is summed over the generators of the gauge group G, and the fermion multiplet ψ belongs to an irreducible representation r of G. The field strength is

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu, \tag{16.2}$$

where f^{abc} are the structure constants of G. The covariant derivative is defined in terms of the representation matrices t^a_r by

$$D_\mu = \partial_\mu - igA^a_\mu t^a_r. \tag{16.3}$$

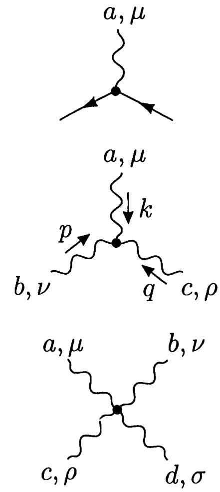
From now on we will drop the subscript r except where it is needed for clarity.

The Feynman rules for this Lagrangian can be derived from a functional integral over the fields ψ, $\bar{\psi}$, and A^a_μ. Imagine expanding the functional integral in perturbation theory, starting with the free Lagrangian, at $g = 0$. The free theory contains a number of free fermions equal to the dimension $d(r)$ of the representation r, and a number of free vector bosons equal to the number $d(G)$ of generators of G. Using the methods of Section 9.5, it is straightforward to derive the fermion propagator

$$\langle \psi_{i\alpha}(x)\bar{\psi}_{j\beta}(y)\rangle = \int \frac{d^4k}{(2\pi)^4}\left(\frac{i}{\not{k} - m}\right)_{\alpha\beta}\delta_{ij}\, e^{-ik\cdot(x-y)}, \tag{16.4}$$

where α, β are Dirac indices and i, j are indices of the symmetry group: $i, j = 1, \ldots, d(r)$. In analogy with electrodynamics, we would guess that the propagator of the vector fields is

$$\langle A^a_\mu(x)A^b_\nu(y)\rangle = \int \frac{d^4k}{(2\pi)^4}\left(\frac{-ig_{\mu\nu}}{k^2}\right)\delta^{ab}\, e^{-ik\cdot(x-y)}, \tag{16.5}$$



$$= \quad ig\gamma^\mu t^a$$

$$= \quad gf^{abc}\big[g^{\mu\nu}(k-p)^\rho + g^{\nu\rho}(p-q)^\mu + g^{\rho\mu}(q-k)^\nu\big]$$

$$= \quad -ig^2\big[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})\big]$$

Figure 16.1. Feynman rules for fermion and gauge boson vertices of a non-Abelian gauge theory.

with $a, b = 1, \ldots, d(G)$. We will derive this formula in the next section.

To find the vertices, we write out the nonlinear terms in (16.1). If $\mathcal{L}_0$ is the free field Lagrangian, then

$$\begin{aligned}\mathcal{L} = \mathcal{L}_0 + gA^a_\lambda\bar{\psi}\gamma^\lambda t^a\psi - gf^{abc}(\partial_\kappa A^a_\lambda)A^{\kappa b}A^{\lambda c} \\ - \tfrac{1}{4}g^2(f^{eab}A^a_\kappa A^b_\lambda)(f^{ecd}A^{\kappa c}A^{\lambda d}).\end{aligned} \tag{16.6}$$

The first of the three nonlinear terms gives the fermion-gauge boson vertex

$$ig\gamma^\mu t^a; \tag{16.7}$$

this is a matrix that acts on the Dirac and gauge indices of the fermions. The second nonlinear term leads to a three gauge boson vertex. To work out this vertex, we first choose a definite convention for the external momenta and Lorentz and gauge indices. A suitable convention is shown in Fig. 16.1, with all momenta pointing inward. Consider first contracting the external gauge particle with momentum k to the first factor of A^a_μ, the gauge particle with momentum p to the second, and the gauge particle of momentum q to the third. The derivative contributes a factor $(-ik_\kappa)$ if the momentum points into the diagram. Then this contribution is

$$-igf^{abc}(-ik^\nu)g^{\mu\rho}. \tag{16.8}$$

In all, there are 3! possible contractions, which alternate in sign according to the total antisymmetry of f^{abc}. The sum of these is exhibited in Fig. 16.1. Finally, the last term of (16.6) leads to a four gauge boson vertex. Following the conventions of Fig. 16.1, one possible contraction gives the contribution

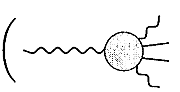
$$-ig^2f^{eab}f^{ecd}g^{\mu\rho}g^{\nu\sigma}. \tag{16.9}$$

There are 4! possible contractions, of which sets of 4 are equal to one another. The sum of these contributions is shown in Fig. 16.1.

Notice that all of these vertices involve the same coupling constant g. We derived the vertices, and thus the equality of the coupling constants, as a part of our construction of the Lagrangian from the principle of non-Abelian gauge invariance. However, it is also possible to see the need for this equality *a posteriori*, from the properties of Feynman amplitudes.

Equality of Coupling Constants

One property that we expect from Feynman amplitudes in non-Abelian gauge theories is that they should satisfy Ward identities similar to those of QED. These Ward identities express the conservation of the symmetry currents, which follows already from the global symmetry of the theory. In QED, the simplest form of the Ward identity was obtained by putting external electrons and positrons on shell. In non-Abelian gauge theories, the gauge bosons also carry charge and so these must also be put on shell to remove contact terms. With all external particles on shell, the amplitude for production of a gauge boson should obey

$k^\mu\Big($  $\Big) = 0.$ (16.10)

This identity is not only an indication of the local gauge symmetry, but is physically important in its own right. Like the photon, the non-Abelian gauge boson has only two physical polarization states. In QED, the on-shell Ward identity expressed the fact that the orthogonal, unphysical polarization states are not produced in scattering processes. The on-shell Ward identity will play a similar role in the non-Abelian case.

Let us check the Ward identity in a simple case, the lowest-order diagrams contributing to fermion-antifermion annihilation into a pair of gauge bosons. In order g^2, there are three diagrams, shown in Fig. 16.2. The first two diagrams are similar to the QED diagrams that we studied in Section 5.5; they sum to

$$\begin{aligned} i\mathcal{M}^{\mu\nu}_{1,2}\epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2) = (ig)^2\bar{v}(p_+)\Big\{&\gamma^\mu t^a \frac{i}{\not{p}-\not{k}_2-m}\gamma^\nu t^b \\ &+ \gamma^\nu t^b \frac{i}{\not{k}_2-\not{p}_+-m}\gamma^\mu t^a\Big\}u(p)\,\epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2). \end{aligned} \tag{16.11}$$

The vectors $\epsilon(k_i)$ are the gauge boson polarization vectors; for physical polarizations, these satisfy $k_i^\mu \epsilon_\mu(k_i) = 0$. To check the Ward identity (16.10), we

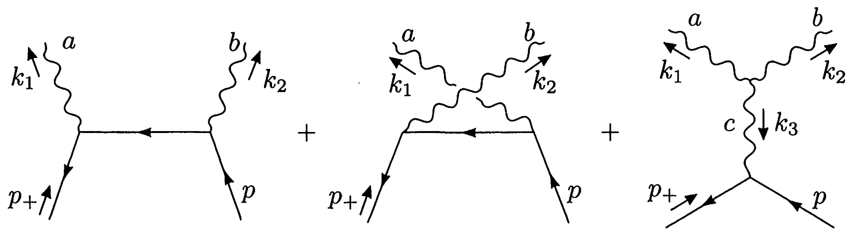


Figure 16.2. Diagrams contributing to fermion-antifermion annihilation to two gauge bosons.

replace $\epsilon_\nu^*(k_2)$ in (16.11) by $k_{2\nu}$. This gives

$$i\mathcal{M}_{1,2}^{\mu\nu}\epsilon_{1\mu}^* k_{2\nu} = (ig)^2\bar{v}(p_+)\Big\{\gamma^\mu t^a \frac{i}{\not{p}-\not{k}_2-m}\not{k}_2 t^b + \not{k}_2 t^b \frac{i}{\not{k}_2-\not{p}_+-m}\gamma^\mu t^a\Big\}u(p)\,\epsilon_{1\mu}^*. \tag{16.12}$$

Since

$$(\not{p}-m)u(p)=0 \qquad \text{and} \qquad \bar{v}(p_+)(-\not{p}_+-m)=0, \tag{16.13}$$

we can add these quantities to $\not{k}_2$ in the first and second terms of (16.12), to cancel the denominators. This gives

$$i\mathcal{M}_{1,2}^{\mu\nu}\epsilon_{1\mu}^* k_{2\nu} = (ig)^2\bar{v}(p_+)\Big\{-i\gamma^\mu[t^a,t^b]\Big\}u(p)\,\epsilon_{1\mu}^*. \tag{16.14}$$

In the Abelian case, this expression would vanish. In the non-Abelian case, however, the residual term is nonzero and depends on the commutator of gauge group generators:

$$i\mathcal{M}_{1,2}^{\mu\nu}\epsilon_{1\mu}^* k_{2\nu} = -g^2\bar{v}(p_+)\gamma^\mu u(p)\,\epsilon_{1\mu}^*\cdot f^{abc}t^c. \tag{16.15}$$

We need to find another contribution to cancel this term. Notice, however, that this term has the group index structure of a fermion-gauge boson vertex ($g\gamma^\mu t^c$) multiplied by a three gauge boson vertex (gf^{abc}). This is just the structure of the third diagram in Fig. 16.2.

To check that the cancellation works, let us evaluate the third diagram:

$$\begin{aligned} i\mathcal{M}_3^{\mu\nu}\epsilon_{1\mu}^*\epsilon_{2\nu}^* &= ig\bar{v}(p_+)\gamma_\rho t^c u(p)\,\frac{-i}{k_3^2}\,\epsilon_\mu^*(k_1)\epsilon_\nu^*(k_2)\\ &\qquad\times gf^{abc}\big[g^{\mu\nu}(k_2-k_1)^\rho + g^{\nu\rho}(k_3-k_2)^\mu + g^{\rho\mu}(k_1-k_3)^\nu\big], \end{aligned}$$

with $k_3=-k_1-k_2$. If we replace $\epsilon_\nu^*(k_2)$ with $k_{2\nu}$, then eliminate k_2 using momentum conservation, the expression in brackets simplifies as follows:

$$\begin{aligned} &\epsilon_\nu^*(k_2)\big[g^{\mu\nu}(k_2-k_1)^\rho + g^{\nu\rho}(k_3-k_2)^\mu + g^{\rho\mu}(k_1-k_3)^\nu\big]\\ &\quad\to k_2^\mu(k_2-k_1)^\rho + k_2^\rho(k_3-k_2)^\mu + g^{\rho\mu}(k_1-k_3)\cdot k_2\\ &\quad= g^{\rho\mu}k_3^2 - k_3^\rho k_3^\mu - g^{\rho\mu}k_1^2 + k_1^\rho k_1^\mu. \end{aligned} \tag{16.16}$$

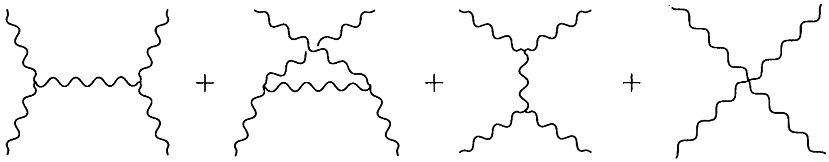


Figure 16.3. Diagrams contributing to gauge boson-gauge boson scattering.

Let us assume that the other gauge boson, with momentum k_1, is on shell ($k_1^2 = 0$), and that it has transverse polarization ($k_1^\mu \epsilon_\mu(k_1) = 0$). Then the third and fourth terms in the last line vanish. Furthermore, the term $k_3^\rho k_3^\mu$ vanishes when it is contracted with the fermion current. In the remaining term, the factor k_3^2 cancels the gauge boson propagator, and we are left with

$$i\mathcal{M}_3^{\mu\nu}\epsilon_{1\mu}^* k_{2\nu} = +g^2 \bar{v}(p_+)\gamma^\mu u(p)\, \epsilon_{1\mu}^* \cdot f^{abc}t^c, \tag{16.17}$$

which precisely cancels (16.15).

Notice that this cancellation takes place only if the value of the coupling constant in the three-boson vertex is identical to that in the fermion-boson vertex. In a similar way, the Ward identity cannot be satisfied among the diagrams for boson-boson scattering, shown in Fig. 16.3, unless the coupling constant g in the four-boson vertex is identical to that in the three-boson vertex. Thus, the coupling constants of all three nonlinear terms in the Yang-Mills Lagrangian *must* be equal in order to preserve the Ward identity and avoid the production of bosons with unphysical polarization states. Conversely, the non-Abelian gauge symmetry guarantees that these couplings *are* equal. The symmetry thus accomplishes exactly what we hoped it would in our discussion at the beginning of Chapter 15, giving us a consistent theory of physical vector particle interactions.

A Flaw in the Argument

The preceding argument has one serious deficiency. At the final stage, we needed to assume that the second gauge boson was transverse. However, one might have expected that this information would come out of the argument rather than having to be put in. In QED, the Feynman diagrams predict that, when an electron and a positron annihilate to form two photons, only the physical transverse polarization states of the photons are produced. Amplitudes to produce other photon polarizations cancel each other to yield zero, as we saw in Eq. (5.80). This statement is not true for the non-Abelian gauge theory Feynman rules that we have worked with so far.

To state the discrepancy more concretely, we introduce some notation. Let $k^\mu = (k^0, \mathbf{k})$ be a lightlike vector: $k^2 = 0$. Then there are two purely spatial vectors orthogonal to $\mathbf{k}$. If k is the momentum of a vector boson, these are the two transverse polarizations. To construct an orthogonal basis, we must include also the longitudinal polarization state, with polarization vector parallel to $\mathbf{k}$, and the timelike polarization state. It is most convenient to work

with the two lightlike linear combinations of these states, with polarization vectors parallel to the vectors k^μ and $\tilde{k}^\mu = (k^0, -\mathbf{k})$. These two unphysical polarization states of a massless vector particle can be written as follows:

$$\epsilon^+_\mu(k) = \left(\frac{k^0}{\sqrt{2}|\mathbf{k}|}, \frac{\mathbf{k}}{\sqrt{2}|\mathbf{k}|}\right); \qquad \epsilon^-_\mu(k) = \left(\frac{k^0}{\sqrt{2}|\mathbf{k}|}, -\frac{\mathbf{k}}{\sqrt{2}|\mathbf{k}|}\right). \tag{16.18}$$

We will refer to $\epsilon^+(k)$ and $\epsilon^-(k)$ as the *forward* and *backward* lightlike polarization vectors. Denote the two transverse polarization states $\epsilon^T_{i\mu}(k)$, for $i = 1, 2$. These four polarization vectors obey the orthogonality relations

$$\begin{aligned} \epsilon^T_i \cdot \epsilon^{*T}_j = -\delta_{ij}, \qquad \epsilon^+ \cdot \epsilon^T_i = \epsilon^- \cdot \epsilon^T_i = 0, \\ (\epsilon^+)^2 = (\epsilon^-)^2 = 0, \qquad \epsilon^+ \cdot \epsilon^- = 1. \end{aligned} \tag{16.19}$$

They also satisfy the completeness relation

$$g_{\mu\nu} = \epsilon^-_\mu \epsilon^{+*}_\nu + \epsilon^+_\mu \epsilon^{-*}_\nu - \sum_i \epsilon^T_{i\mu} \epsilon^{T*}_{i\nu}. \tag{16.20}$$

Using this notation, we can express concretely the gap in the argument for the Ward identity. The Feynman diagrams of Fig. 16.2 apparently predict that there is a nonzero amplitude to produce a forward-polarized gauge boson together with a backward-polarized gauge boson. For this case, we substitute $\epsilon^{-*}_\mu(k_1)$ and $\epsilon^{+*}_\nu(k_2)$ for the two polarization vectors. Then the term proportional to $k_1^\rho k_1^\mu$ in Eq. (16.16) no longer vanishes; it now yields

$$\begin{aligned} i\mathcal{M} &= ig\bar{v}(p_+)\gamma_\rho t^c u(p)\,\frac{-i}{k_3^2}\,\epsilon^{-*}_\mu(k_1)\cdot\frac{1}{\sqrt{2}|\mathbf{k}_2|}\cdot g f^{abc} k_1^\rho k_1^\mu \\ &= ig\bar{v}(p_+)\gamma_\rho t^c u(p)\,\frac{-i}{k_3^2}\cdot g f^{abc} k_1^\rho \cdot \frac{|\mathbf{k}_1|}{|\mathbf{k}_2|}. \end{aligned} \tag{16.21}$$

Can we simply ignore this totally unphysical process? We are free to do so in calculations of leading-order amplitudes, but the process will come back to haunt us in loop diagrams. Recall from Section 7.3 how the optical theorem (7.49) links the imaginary part of a loop diagram to the square of a corresponding scattering amplitude, obtained by cutting the diagram across the loop. If we apply the optical theorem to the diagram shown in Fig. 16.4, we obtain a paradox. In the gauge boson loop on the left-hand side we can replace the $g_{\mu\nu}$ factors in the propagators with sums over all four polarization vectors (16.20). The theorem thus implies that all four polarizations, even the unphysical ones, should be included for the final-state gauge bosons on the right-hand side. We are faced with a choice of allowing the production of unphysical states or violating the optical theorem. A third alternative, equally unattractive, would be to discard our expression (16.5) for the gauge boson propagator. Clearly, we are missing some crucial element of the quantum-mechanical structure of non-Abelian gauge theories.

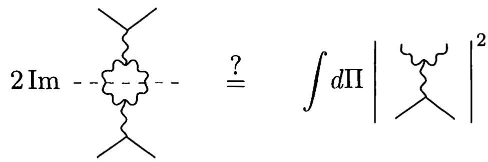


Figure 16.4. A paradox for the optical theorem in gauge theories.

16.2 The Faddeev-Popov Lagrangian

It is not surprising that we have found a problem with our Feynman rules for non-Abelian gauge theories, since we were not very careful in deriving them. In particular, we did not actually derive expression (16.5) for the gauge field propagator. In this section we will remedy this by going through a formal derivation of this expression. We will find that, although expression (16.5) is indeed correct, it is incomplete: It must be supplemented by additional rules of a completely new type.

To define the functional integral for a theory with non-Abelian gauge invariance, we will use the Faddeev-Popov method, as introduced in Section 9.4 to quantize the electromagnetic field. Our present discussion will follow Section 9.4 closely. However, as we have by now come to expect, the case of non-Abelian local symmetry brings with it new tricks and surprises.

First consider the quantization of the pure gauge theory, without fermions. To derive the Feynman rules, we must define the functional integral

$$\int \mathcal{D}A \, \exp\left[i\int d^4x\left(-\tfrac{1}{4}(F^a_{\mu\nu})^2\right)\right]. \tag{16.22}$$

As in the Abelian case, the Lagrangian is unchanged along the infinite number of directions in the space of field configurations corresponding to local gauge transformations. To compute the functional integral we must factor out the integrations along these directions, constraining the remaining integral to a much smaller space.

As in electrodynamics, we will constrain the gauge directions by applying a gauge-fixing condition $G(A) = 0$ at each point x. Following Faddeev and Popov, we can introduce this constraint by inserting into the functional integral the identity (9.53):

$$1 = \int \mathcal{D}\alpha(x)\, \delta\big(G(A^\alpha)\big)\, \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right). \tag{16.23}$$

Here A^α is the gauge field A transformed through a finite gauge transformation as in (15.47):

$$(A^\alpha)^a_\mu t^a = e^{i\alpha^a t^a}\left[A^b_\mu t^b + \frac{i}{g}\partial_\mu\right]e^{-i\alpha^c t^c}. \tag{16.24}$$

In evaluating the determinant, the infinitesimal form of this transformation will be more useful:

$$(A^\alpha)^a_\mu = A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c = A^a_\mu + \frac{1}{g} D_\mu \alpha^a, \tag{16.25}$$

where D_μ is the covariant derivative (15.86) acting on a field in the adjoint representation. Note that, as long as the gauge-fixing function $G(A)$ is linear, the functional derivative $\delta G(A^\alpha)/\delta\alpha$ is independent of α.

Since the Lagrangian is gauge invariant, we can replace A by A^α in the exponential of (16.22). Then, as in the Abelian case, we can interchange the order of the functional integrals over A and α, and then change variables in the inner integral from A to $A' = A^\alpha$. The transformation (16.24) looks more complicated than in the Abelian case, but it is nothing more than a linear shift of the A^a_μ, followed by a unitary rotation of the various components of the symmetry multiplet $A^a_\mu(x)$ at each point. Both of these operations preserve the measure

$$\mathcal{D}A = \prod_x \prod_{a,\mu} dA^a_\mu. \tag{16.26}$$

Thus $\mathcal{D}A = \mathcal{D}A'$, under the integral over α. Just as in the Abelian case, the integral over gauge motions α can be factored out of the functional integral into an overall normalization, leaving us with

$$\int \mathcal{D}A\, e^{iS[A]} = \left(\int \mathcal{D}\alpha\right) \int \mathcal{D}A\, e^{iS[A]}\, \delta\big(G(A)\big)\, \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right). \tag{16.27}$$

This normalization factor cancels in the computation of correlation functions of gauge-invariant operators.

From this point, the derivation of the gauge boson propagator proceeds as for the photon propagator. We choose the generalized Lorentz gauge condition

$$G(A) = \partial^\mu A^a_\mu(x) - \omega^a(x), \tag{16.28}$$

with a Gaussian weight for ω^a as in Eq. (9.56). The manipulations of Section 9.4 then lead to the class of gauge field propagators

$$\big\langle A^a_\mu(x) A^b_\nu(y)\big\rangle = \int \frac{d^4k}{(2\pi)^4} \frac{-i}{k^2 + i\epsilon}\left(g_{\mu\nu} - (1-\xi)\frac{k_\mu k_\nu}{k^2}\right)\delta^{ab} e^{-ik\cdot(x-y)}, \tag{16.29}$$

with a freely adjustable gauge parameter ξ. Our guess (16.5) corresponds to the choice $\xi = 1$, called the *Feynman-'t Hooft gauge.*

So far, this whole derivation parallels the case of electrodynamics. Here, however, there is one more nontrivial ingredient. In QED, the determinant in Eq. (16.23) was independent of A, so this quantity could be treated as just another contribution to the normalization factor. In the non-Abelian case this is no longer true. Using the infinitesimal form (16.25) of the gauge transformation, we can evaluate

$$\frac{\delta G(A^\alpha)}{\delta\alpha} = \frac{1}{g}\partial^\mu D_\mu, \tag{16.30}$$

acting on a field in the adjoint representation; this operator depends on A. The functional determinant of (16.30) thus contributes new terms to the Lagrangian.

Faddeev and Popov chose to represent this determinant as a functional integral over a new set of anticommuting fields belonging to the adjoint representation:

$$\det\left(\frac{1}{g}\partial^\mu D_\mu\right) = \int \mathcal{D}c\mathcal{D}\bar{c}\, \exp\left[i\int d^4x\, \bar{c}(-\partial^\mu D_\mu)c\right]. \tag{16.31}$$

We derived this formal identity in Eq. (9.69), using our rules for fermionic functional integrals. (The factor of $1/g$ is absorbed into the normalization of the fields c and $\bar{c}$.) But to give the correct identity, c and $\bar{c}$ must be anticommuting fields that are scalars under Lorentz transformations. The quantum excitations of these fields have the wrong relation between spin and statistics to be physical particles. However, we can nevertheless treat these excitations as additional particles in the computation of Feynman diagrams. These new fields and their particle excitations are called *Faddeev-Popov ghosts.*

If we temporarily suppress our curiosity about the physical interpretation of the ghosts, we can work out their Feynman rules. We write the ghost Lagrangian more explicitly as

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a\left(-\partial^2\delta^{ac} - g\partial^\mu f^{abc}A^b_\mu\right)c^c. \tag{16.32}$$

The first term gives a ghost propagator,

$$\left\langle c^a(x)\bar{c}^b(y)\right\rangle = \int \frac{d^4k}{(2\pi)^4}\frac{i}{k^2}\delta^{ab}e^{-ik\cdot(x-y)}. \tag{16.33}$$

In a diagram, this propagator carries an arrow that shows the flow of ghost number, as in Fig. 16.5. In the interaction term of (16.32), the derivative stands to the left of the gauge field; this implies that this derivative is evaluated with the momentum coming out of the vertex along the ghost line. The explicit Feynman rule is shown in Fig. 16.5. As with the other vertices we have encountered, the coupling constant g that appears in this vertex must be equal to the coupling constant g in the three-boson vertex in order to avoid upsetting the Ward identities.

There are no further subtleties in the construction of the perturbation theory for non-Abelian gauge theories. In particular, it is straightforward to include fermions. The final Lagrangian, including all of the effects of Faddeev-Popov gauge fixing, is

$$\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2 - \frac{1}{2\xi}(\partial^\mu A^a_\mu)^2 + \bar{\psi}(i\not{D} - m)\psi + \bar{c}^a(-\partial^\mu D^{ac}_\mu)c^c. \tag{16.34}$$

This Lagrangian leads to the propagator (16.29), and to the set of Feynman rules for vertices shown in Figs. 16.1 and 16.5.

The argument we have just completed suffices to derive the Feynman diagram expansion of any correlation function of gauge-invariant operators in a non-Abelian gauge theory. At the end of Section 9.4, we explained that the

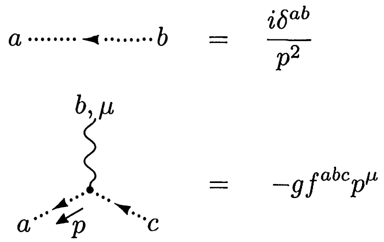


Figure 16.5. Feynman rules for Faddeev-Popov ghosts.

Faddeev-Popov gauge-fixing technique also gives the correct gauge-invariant expressions for S-matrix elements. This remains true in the non-Abelian case. However, the argument given in Section 9.4 relied upon the cancellation in QED of the emission probabilities for timelike and longitudinal photons, and we have already found that this cancellation does not go through in the non-Abelian case. In Section 16.4 we will construct a more sophisticated argument, in which the Faddeev-Popov ghosts play an essential role, that will correctly generalize our previous argument to non-Abelian gauge theories.

16.3 Ghosts and Unitarity

We might now ask whether the new ingredients that we found in the previous section, the Faddeev-Popov ghosts, can resolve the paradox that we encountered at the end of Section 16.1. There we saw that the first diagram in Fig. 16.6 contains a nonzero contribution to its imaginary part that does not correspond to a possible final state with physical gauge boson polarizations. We will now compute this contribution more carefully. We must then add a new potential contribution from the ghosts, shown as the second diagram in Fig. 16.6.

Let us call the amplitude for fermion-fermion annihilation into gauge bosons, which we studied in Section 16.1,

$$i\mathcal{M}^{\mu\nu}\epsilon^*_\mu(k_1)\epsilon^*_\nu(k_2); \tag{16.35}$$

the amplitude for two gauge bosons to convert to a fermion-antifermion pair will be, correspondingly, $\mathcal{M}'$. Then, following the Cutkosky rules of Section 7.3, we find the imaginary part of the first diagram in Fig. 16.6 by replacing the cut gauge boson propagator with momentum k_i by

$$-ig_{\mu\nu}\cdot(-2\pi i)\delta(k_i^2). \tag{16.36}$$

Replacing both propagators gives two delta functions, turning the four-dimensional integrals over the gauge boson momenta into three-dimensional phase space integrals, as in the example in Section 7.3. We are thus left with the expression

$$\tfrac{1}{2}(i\mathcal{M}^{\mu\nu})g_{\mu\rho}g_{\nu\sigma}(i\mathcal{M}'^{\rho\sigma}), \tag{16.37}$$

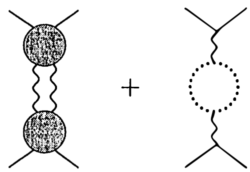


Figure 16.6. The diagram on the left, in which each circle represents the sum of the three contributions of Fig. 16.2, gives a possible problem for the optical theorem. The ghost diagram on the right cancels the anomalous terms.

integrated over the phase space of two massless particles. The factor 1/2 is a symmetry factor for the Feynman diagram or, equivalently, a correction to the phase space integral for identical particles.

Now introduce the representation (16.20) for $g_{\mu\rho}$ and $g_{\nu\sigma}$. The pieces that involve only transverse polarizations correspond to the expected imaginary parts necessary to satisfy the optical theorem. We need not consider these terms further. The cross terms between physical and unphysical polarizations vanish: We showed in Section 16.1 that

$$i\mathcal{M}^{\mu\nu}\epsilon_\mu^{T*}(k_1)\epsilon_\nu^{+*}(k_2) = 0. \tag{16.38}$$

The same identity holds if $\mathcal{M}$ is replaced by $\mathcal{M}'$, and if ϵ^+ is replaced by ϵ^-. Furthermore, the amplitude vanishes if both polarization vectors are forward or both are backward. The only surviving terms are the cross terms between forward and backward polarization, which yield the expression

$$\tfrac{1}{2}\left[(i\mathcal{M}^{\mu\nu}\epsilon_\mu^{-*}\epsilon_\nu^{+*})(i\mathcal{M}'^{\rho\sigma}\epsilon_\rho^{+}\epsilon_\sigma^{-}) + (i\mathcal{M}^{\mu\nu}\epsilon_\mu^{+*}\epsilon_\nu^{-*})(i\mathcal{M}'^{\rho\sigma}\epsilon_\rho^{-}\epsilon_\sigma^{+})\right], \tag{16.39}$$

integrated over phase space. We worked out the value of the first factor in Eq. (16.21), and the contraction with $\mathcal{M}'$ is very similar. Substituting these results, expression (16.39) becomes

$$\begin{aligned}\frac{1}{2}&\left(ig\bar{v}(p_+)\gamma_\mu t^c u(p)\cdot\frac{-i}{(k_1+k_2)^2}\cdot gf^{abc}k_1^\mu\right)\\ &\times\left(ig\bar{u}(p')\gamma_\rho t^d v(p'_+)\cdot\frac{-i}{(k_1+k_2)^2}\cdot gf^{abd}(-k_2)^\rho\right)+(k_1\leftrightarrow k_2).\end{aligned}\tag{16.40}$$

Using the identity

$$\bar{v}(p_+)\gamma_\mu(k_1+k_2)^\mu u(p) = \bar{v}(p_+)\gamma_\mu(p+p_+)^\mu u(p) = 0, \tag{16.41}$$

we see that the two terms added in (16.40) are equal.

Now add the contribution from the Faddeev-Popov ghosts. Using the Feynman rules in Fig. 16.5, we can assemble the amplitude for fermion-antifermion annihilation into a pair of ghosts:

$$i\mathcal{M}_{\text{ghost}} = ig\bar{v}(p_+)\gamma_\mu t^c u(p)\cdot\frac{-i}{(k_1+k_2)^2}\cdot gf^{abc}k_1^\mu. \tag{16.42}$$

This is precisely the first half of expression (16.40). Similarly, the amplitude for the ghost-antighost pair to annihilate into fermions is equal to the second half of (16.40). Finally, since Faddeev-Popov ghost fields anticommute, we must supply a factor of -1 for each ghost loop. Thus the ghost contribution exactly cancels the contribution of unphysical gauge boson polarizations to the Cutkosky cut of the diagrams in Fig. 16.6.

This example illustrates a general physical interpretation of Faddeev-Popov ghosts. These "particles" serve as negative degrees of freedom to cancel the effects of the unphysical timelike and longitudinal polarization states of the gauge bosons. The simplest effect of the ghosts can already be seen from the determinants that appear when one integrates over the gauge and ghost fields in the Faddeev-Popov Lagrangian (16.34). In a general dimension d, working in Feynman gauge and at zero coupling for simplicity, the functional integral over the gauge and ghost fields in (16.34) yields

$$\left(\det[-\partial^2]\right)^{-d/2} \cdot \left(\det[-\partial^2]\right)^{+1}. \tag{16.43}$$

The second determinant, which appears with a positive exponent because the ghost fields anticommute, cancels the contribution to the first determinant of two components of the field A_μ. This physical effect was illustrated, using the language of Section 9.4, in Problem 9.2.

16.4 BRST Symmetry

To show how this cancellation extends to the complete interacting theory, Becchi, Rouet, Stora, and Tyutin introduced as a beautiful formal tool a new symmetry of the gauge-fixed Lagrangian (16.34), which involves the ghost in an essential way.* This *BRST symmetry* has a continuous parameter that is an anticommuting number. To write the symmetry in its simplest form, let us rewrite the Faddeev-Popov Lagrangian by introducing a new (commuting) scalar field B^a:

$$\mathcal{L} = -\frac{1}{4}(F^a_{\mu\nu})^2 + \bar{\psi}(i\not{D} - m)\psi + \frac{\xi}{2}(B^a)^2 + B^a\partial^\mu A^a_\mu + \bar{c}^a(-\partial^\mu D^{ac}_\mu)c^c. \tag{16.44}$$

The new field B^a has a quadratic term without derivatives, so it is not a normal propagating field. The functional integral over B^a can be done by completing the square in (16.44); this procedure brings us back precisely to Eq. (16.34). A field of this type, which appears in the functional integral but has no independent dynamics, is called an *auxiliary field.*

*C. Becchi, A. Rouet, and R. Stora, *Ann. Phys.* **98**, 287 (1976); I. V. Tyutin, Lebedev Institute preprint (1975, unpublished); M. Z. Iofa and I. V. Tyutin, *Theor. Math. Phys.* **27**, 316 (1976).

Now let ϵ be an infinitesimal anticommuting parameter, and consider the following infinitesimal transformation of the fields in (16.44):

$$\begin{aligned} \delta A^a_\mu &= \epsilon D^{ac}_\mu c^c \\ \delta\psi &= ig\epsilon c^a t^a \psi \\ \delta c^a &= -\tfrac{1}{2} g\epsilon f^{abc} c^b c^c \\ \delta \bar{c}^a &= \epsilon B^a \\ \delta B^a &= 0. \end{aligned} \tag{16.45}$$

The transformation of the fields A^a_μ and ψ is a local gauge transformation whose parameter is proportional to the ghost field: $\alpha^a(x) = g\epsilon c^a(x)$. Thus, the first two terms of (16.44) are invariant to (16.45). The third term is trivially invariant. The transformation of A^a_μ in the fourth term cancels the transformation of $\bar{c}^a$ in the last term. Finally, we must examine the transformation of the last ingredient in (16.44):

$$\begin{aligned} \delta(D^{ac}_\mu c^c) &= D^{ac}_\mu \delta c^c + g f^{abc} \delta A^b_\mu c^c \\ &= -\tfrac{1}{2} g\epsilon \partial_\mu (f^{abc} c^b c^c) - \tfrac{1}{2} g^2 \epsilon f^{abc} f^{cde} A^b_\mu c^d c^e \\ &\qquad + g\epsilon f^{abc} (\partial_\mu c^b) c^c + g^2 \epsilon f^{abc} f^{bde} A^d_\mu c^e c^c. \end{aligned} \tag{16.46}$$

The two terms of order g manifestly cancel. By using the anticommuting nature of the ghost fields and exchanging the names of indices, we can write the remaining two terms as

$$-\tfrac{1}{2} \epsilon g^2 f^{abc} f^{cde} \big(A^b_\mu c^d c^e + A^d_\mu c^e c^b + A^e_\mu c^b c^d \big), \tag{16.47}$$

which vanishes by the Jacobi identity (15.70). Apparently, the BRST transformation (16.45) is a global symmetry of the gauge-fixed Lagrangian (16.44), for any value of the gauge parameter ξ.

The BRST transformation has one more remarkable feature, which is a natural consequence of its anticommuting nature. Let $Q\phi$ be the BRST transformation of the field ϕ: $\delta\phi = \epsilon Q\phi$. For example, $QA^a_\mu = D^{ac}_\mu c^c$. Then, for any field, the BRST variation of $Q\phi$ vanishes:

$$Q^2 \phi = 0. \tag{16.48}$$

The vanishing of (16.46) proves this identity for the second BRST variation of the gauge field. For the ghost field,

$$Q^2 c^a = \tfrac{1}{2} g^2 f^{abc} f^{bde} c^c c^d c^e, \tag{16.49}$$

which vanishes by the Jacobi identity. It is straightforward to check that the second BRST variations of the other fields in (16.44) also vanish.

To describe the implications of identity (16.48), we now consider studying

the effective theory (16.44) in the Hamiltonian picture after canonical quantization. Because the Lagrangian has the continuous symmetry (16.45), the theory will have a conserved current, and the integral of the time component of this current will be a conserved charge Q that commutes with H. The action of Q on field configurations will be just that described in the previous paragraph. The relation (16.48) is equivalent to the operator identity

$$Q^2 = 0. \tag{16.50}$$

We say that the BRST operator Q is *nilpotent.*

A nilpotent operator that commutes with H divides the eigenstates of H into three subspaces. Many eigenstates of H must be annihilated by Q so that (16.50) can be satisfied. Let $\mathcal{H}_1$ be the subspace of states that are *not* annihilated by Q. Let $\mathcal{H}_2$ be the subspace of states of the form

$$|\psi_2\rangle = Q\,|\psi_1\rangle\,, \tag{16.51}$$

where $|\psi_1\rangle$ is in $\mathcal{H}_1$. According to (16.50), acting Q again on these states gives zero. Finally, let $\mathcal{H}_0$ be the subspace of states $|\psi_0\rangle$ that satisfy $Q\,|\psi_0\rangle = 0$ but that cannot be written in the form (16.51). The subspace $\mathcal{H}_2$ is quite peculiar, because any two states in this subspace have zero inner product:

$$\langle\psi_{2a}|\psi_{2b}\rangle = \langle\psi_{1a}|\,Q\,|\psi_{2b}\rangle = 0 \tag{16.52}$$

by (16.50). By the same argument, the states of $\mathcal{H}_2$ have zero inner product with the states of $\mathcal{H}_0$.

These considerations seem extremely abstract, but they have a direct physical correspondence.† To see this, consider single-particle states of the non-Abelian gauge theory in the limit of zero coupling. According to the transformation (16.45), Q converts the forward component of A^a_μ to a ghost field; equivalently, Q converts a single forward-polarized gauge boson to a ghost. At $g = 0$, Q annihilates the one-ghost state. At the same time, Q converts the antighost state to a quantum of B^a. To identify this state, note that the Lagrangian (16.44) implies the classical field equation

$$\xi B^a = -\partial^\mu A^a_\mu. \tag{16.53}$$

Thus the quanta of the field B^a are those quanta of A^a_μ with polarization vectors such that $k^\mu\epsilon_\mu(k) \neq 0$; these are the backward-polarized gauge bosons.

We have now seen that, among the single-particle states of the gauge theory, forward gauge bosons and antighosts belong to $\mathcal{H}_1$, ghosts and backward gauge bosons belong to $\mathcal{H}_2$, and transverse gauge bosons belong to $\mathcal{H}_0$. More generally, it can be shown that asymptotic states containing ghosts,

†The following argument is presented only at an intuitive level. For a rigorous discussion, see T. Kugo and I. Ojima, *Suppl. Prog. Theor. Phys.* **66**, 1 (1979).

antighosts, or gauge bosons of unphysical polarization always belong to $\mathcal{H}_1$ or $\mathcal{H}_2$, while the asymptotic states in $\mathcal{H}_0$ are those with only transversely polarized gauge bosons. The BRST operator thus gives a precise relation between the unphysical gauge boson polarization states and the ghosts and antighosts as positive and negative degrees of freedom.

In Section 9.4, we argued that the Faddeev-Popov prescription gave the correct, gauge-invariant result for a certain subclass of S-matrix elements, from which we could compute the physical scattering cross sections of transversely polarized gauge bosons. These S-matrix elements were constructed by putting operators in the far past to create transversely polarized gauge bosons, adiabatically turning on the gauge coupling, adiabatically turning off the gauge coupling, and then placing operators in the far future to annihilate gauge bosons with transverse polarization. However, this argument had a possible problem: If the states created as collections of transversely polarized bosons in the far past could evolve into states that contained gauge bosons of other polarizations in the far future, the S-matrix projected between transverse gauge boson states would not be unitary. This problem would also lead to the technical problem discussed in the previous section: The Cutkosky cuts of diagrams contributing to S-matrix elements would have nonzero contributions from unphysical polarizations. In Section 9.4, we used an argument special to the Abelian case to show that these problems do not arise in QED. In the non-Abelian case, the removal of unphysical gauge boson polarizations is more subtle, and we have seen that it involves the ghosts in an essential way. To resolve this subtle problem, we apply the principle of BRST symmetry.

Let $|A;\mathrm{tr}\rangle$ be an external state that contains no ghosts or antighosts and only gauge bosons with transverse polarization. We wish to show that the S-matrix projected onto such states is unitary:

$$\sum_C \langle A;\mathrm{tr}|\, S^\dagger \,|C;\mathrm{tr}\rangle \, \langle C;\mathrm{tr}|\, S \,|B;\mathrm{tr}\rangle = \langle A;\mathrm{tr}|\, \mathbf{1} \,|B;\mathrm{tr}\rangle \,. \tag{16.54}$$

As we explained above, the physical states $|A;\mathrm{tr}\rangle$ belong to—and, in fact, span—the subspace $\mathcal{H}_0$ defined by the BRST operator. In particular, all of these states are annihilated by Q. Since Q commutes with the Hamiltonian, the time evolution of any such state must also produce a state annihilated by Q. Thus,

$$Q \cdot S\,|A;\mathrm{tr}\rangle = 0. \tag{16.55}$$

This implies that the states $S\,|A;\mathrm{tr}\rangle$ must be linear combinations of states in $\mathcal{H}_0$ and $\mathcal{H}_2$. However, states in $\mathcal{H}_2$ have zero inner product with one another and with states in $\mathcal{H}_0$. Thus the inner product of any two states of the form $S\,|A;\mathrm{tr}\rangle$ comes only from the overlap of the components in $\mathcal{H}_0$, so we can write

$$\langle A;\mathrm{tr}|\, S^\dagger \cdot S \,|B;\mathrm{tr}\rangle = \sum_C \langle A;\mathrm{tr}|\, S^\dagger \,|C;\mathrm{tr}\rangle \, \langle C;\mathrm{tr}|\, S \,|B;\mathrm{tr}\rangle \,. \tag{16.56}$$

Since the full S-matrix is unitary, this relation implies that the restricted S-matrix is also unitary, Eq. (16.54). In addition, (16.56) implies that the sum of the Cutkosky cuts of diagrams contributing to the S-matrix in a given order is equal to the sum of the cuts involving transverse gauge bosons only. Thus, the cancellation between diagrams that produce pairs of gauge bosons with unphysical polarizations and those that produce ghosts is a general property that persists to all orders in perturbation theory.

Since the BRST transformation generates a continuous symmetry, it generates a set of Ward identities. These identities are similar in structure to the Ward identities of the non-Abelian gauge symmetry, since the BRST symmetry contains a gauge transformation whose parameter is the ghost field. However, the identities that follow from BRST symmetry are simpler. We will not study the Ward identities of non-Abelian gauge theory further in this book. However, when one discusses the renormalization of gauge theories at a higher level, the central identities among renormalization constants that follow from the Ward identities are most easily derived using the BRST symmetry.‡

16.5 One-Loop Divergences of Non-Abelian Gauge Theory

Now that we have discussed the general properties of tree-level diagrams in non-Abelian gauge theories, we turn our attention to diagrams with loops. As always in quantum field theory, some of these loop diagrams will diverge, and we must take care to treat the divergent integrals correctly.

The Lagrangian of a non-Abelian gauge theory (15.39) contains no interactions of dimension higher than 4. Therefore, by the general arguments of Chapter 10, this Lagrangian is renormalizable, in the sense that the divergences can be removed by a finite number of counterterms. However, in non-Abelian gauge theories, as in QED, the gauge symmetries of the theory imply stronger restrictions on the structure of the divergences. In QED, provided that we use a gauge-invariant regulator, there are only four possible divergent coefficients, which are subtracted by the counterterms for the electromagnetic vertex (δ_1), for the electron and photon field strength (δ_2 and δ_3), and for the electron mass (δ_m). In particular, the possibility of a photon mass renormalization is excluded by gauge invariance. Furthermore, the two counterterms δ_1 and δ_2 are equal to one another, and cancel in the evaluation of the electron-photon vertex function, as a consequence of the Ward identity. Non-Abelian gauge symmetries imply similar restrictions on the divergences of Feynman diagrams. In this section, we will illustrate some of these restrictions through examples of one-loop diagrams.

‡An introduction to the Ward identities of the BRST symmetry is given by Taylor (1976).

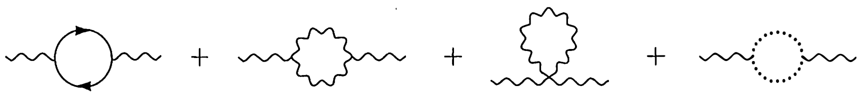
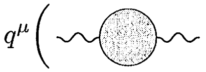


Figure 16.7. Contributions to the gauge boson self-energy in order g^2.

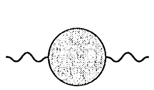
The Gauge Boson Self-Energy

In QED, the strongest constraints of gauge invariance come in the evaluation of the photon self-energy. The Ward identity implies the relation

$$q^{\mu}\Big(\quad\Big) = 0, \qquad (16.57)$$



which in turn implies that the photon self-energy diagrams have the structure

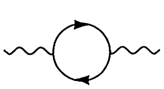


$$= i(q^2 g^{\mu\nu} - q^{\mu}q^{\nu})\Pi(q^2). \qquad (16.58)$$

The only divergence possible is a logarithmically divergent contribution to $\Pi(q^2)$. In non-Abelian gauge theories, (16.57) still holds, so the self-energy again has the Lorentz structure (16.58). However, the cancellations that lead to this structure are more complex. Here we will exhibit these cancellations by computing the gauge boson self-energy in detail at the one-loop level. In order to preserve gauge invariance, we will use dimensional regularization.

The contributions of order g^2 to the gauge boson self-energy are shown in Fig. 16.7. (In addition to these 1PI diagrams, there are three "tadpole" diagrams; but these automatically vanish, as in QED, by the argument given below Eq. (10.5).) The fermion loop diagram can be considered separately from the other diagrams, since in principle we could include any number of fermions in the theory. We will see below that the contributions of the three remaining diagrams interlock in an essential way.

Let us first calculate the fermion loop diagram. The Feynman rule for the vertices in this diagram is identical to the QED Feynman rule, except for the addition of a group matrix t^a that acts on the fermion gauge group indices. The value of this diagram is therefore the same as in QED, Eq. (7.90), multiplied by a trace over group matrices:

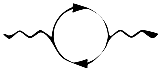


$$= \operatorname{tr}[t^a t^b]\, i(q^2 g^{\mu\nu} - q^{\mu}q^{\nu}) \times \frac{-g^2}{(4\pi)^{d/2}} \int_0^1 dx\, 8x(1-x) \frac{\Gamma(2-\frac{d}{2})}{(m^2 - x(1-x)q^2)^{2-d/2}}.$$

The value of the trace is given by Eq. (15.78): $\operatorname{tr}[t^a t^b] = C(r)\delta^{ab}$. In a theory with several species of fermions, there would be a diagram of this type for each species. We will be mainly interested in the divergent part of this diagram,

which is independent of the fermion mass. If there are n_f species of fermions, all in the same representation r, then the total contribution of fermion loop diagrams takes the form

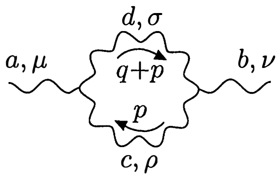
$$\sum_{\text{fermions}} \Bigg(\text{[diagram]} \Bigg)$$



$$= i(q^2 g^{\mu\nu} - q^\mu q^\nu)\delta^{ab}\Big(\frac{-g^2}{(4\pi)^2}\cdot\frac{4}{3}n_f C(r)\Gamma(2-\tfrac{d}{2}) + \cdots\Big). \tag{16.59}$$

Now consider the three diagrams from the pure gauge sector. The contribution of these diagrams depends on the gauge; we will use Feynman-'t Hooft gauge, $\xi = 1$.

Using the three-gauge-boson vertex from Fig. 16.1, we can write the first of the three diagrams as



$$= \frac{1}{2}\int\frac{d^4p}{(2\pi)^4}\frac{-i}{p^2}\frac{-i}{(p+q)^2}g^2 f^{acd}f^{bcd}N^{\mu\nu}, \tag{16.60}$$

where the numerator structure is

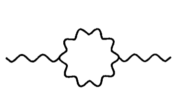
$$N^{\mu\nu} = \big[g^{\mu\rho}(q-p)^\sigma + g^{\rho\sigma}(2p+q)^\mu + g^{\sigma\mu}(-p-2q)^\rho\big] \\ \times \big[\delta^\nu{}_\rho(p-q)_\sigma + g_{\rho\sigma}(-2p-q)^\nu + \delta^\nu{}_\sigma(p+2q)_\rho\big].$$

The overall factor of $1/2$ is a symmetry factor. The contraction of structure constants can be evaluated using Eq. (15.93): $f^{acd}f^{bcd} = C_2(G)\delta^{ab}$.

To simplify the expression further, combine denominators in the standard way:

$$\frac{1}{p^2}\frac{1}{(p+q)^2} = \int_0^1 dx\,\frac{1}{((1-x)p^2 + x(p+q)^2)^2} = \int_0^1 dx\,\frac{1}{(P^2-\Delta)^2}, \tag{16.61}$$

where $P = p + xq$ and $\Delta = -x(1-x)q^2$. Then (16.60) can be rewritten



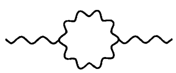
$$= -\frac{g^2}{2}C_2(G)\delta^{ab}\int_0^1 dx\int\frac{d^4P}{(2\pi)^4}\frac{1}{(P^2-\Delta)^2}N^{\mu\nu}.$$

The numerator structure can be simplified by eliminating p in favor of P, discarding terms linear in P^μ (which integrate symmetrically to zero), and

replacing $P^\mu P^\nu$ with $g^{\mu\nu}P^2/d$ (also by symmetry):

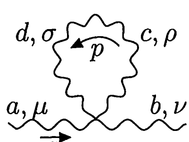
$$\begin{aligned} N^{\mu\nu} = &-g^{\mu\nu}\big[(2q+p)^2 + (q-p)^2\big] - d(q+2p)^\mu(q+2p)^\nu \\ &+ \big[(2q+p)^\mu(q+2p)^\nu + (q-p)^\mu(2q+p)^\nu - (q+2p)^\mu(q-p)^\nu \\ &\quad + (\mu \leftrightarrow \nu)\big] \\ \rightarrow &-g^{\mu\nu}P^2\cdot 6(1-\tfrac{1}{d}) - g^{\mu\nu}q^2\big[(2-x)^2+(1+x)^2\big] \\ &+ q^\mu q^\nu\big[(2-d)(1-2x)^2 + 2(1+x)(2-x)\big]. \end{aligned}$$

The final step in the evaluation is to Wick-rotate and apply the integration formulae (7.85) and (7.86). This brings the diagram into the following form:



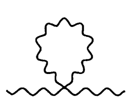
$$\begin{aligned} = &\frac{ig^2}{(4\pi)^{d/2}} C_2(G)\delta^{ab}\int_0^1 dx\,\frac{1}{\Delta^{2-d/2}} \\ &\times\Big(\Gamma(1-\tfrac{d}{2})\,g^{\mu\nu}q^2\,[\tfrac{3}{2}(d-1)x(1-x)] \\ &\quad + \Gamma(2-\tfrac{d}{2})\,g^{\mu\nu}q^2\,[\tfrac{1}{2}(2-x)^2 + \tfrac{1}{2}(1+x)^2] \\ &\quad - \Gamma(2-\tfrac{d}{2})\,q^\mu q^\nu\,[(1-\tfrac{d}{2})(1-2x)^2 + (1+x)(2-x)]\Big). \end{aligned} \tag{16.62}$$

Next consider the diagram with a four-gauge-boson vertex. Using the vertex Feynman rule in Fig. 16.1, we find



$$\begin{aligned} = &\frac{1}{2}\int\frac{d^4p}{(2\pi)^4}\,\frac{-ig_{\rho\sigma}}{p^2}\delta^{cd}\,(-ig^2) \\ &\times\big[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) \\ &+ f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) \\ &+ f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})\big]. \end{aligned} \tag{16.63}$$

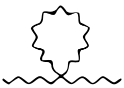
The factor 1/2 in the first line is a symmetry factor. The first combination of structure constants in the vertex factor vanishes by antisymmetry; the second and third can be reduced by the use of Eq. (15.93). We then find simply



$$= -g^2C_2(G)\delta^{ab}\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^2}\cdot g^{\mu\nu}(d-1). \tag{16.64}$$

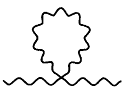
In dimensional regularization, the integral over p gives a pole at $d=2$ but yields zero as $d \rightarrow 4$. We could simply discard this diagram and trust that the pole at $d=2$ is canceled by the other two diagrams. It is instructive, however, and no more difficult, to demonstrate the cancellation explicitly. To do so, we can force the integral to look like that of the previous diagram, multiplying the integrand by 1 in the form $(q+p)^2/(q+p)^2$. We then combine denominators as before, and eliminate p in favor of the shifted variable $P = p + xq$. After

dropping the term linear in P, we obtain



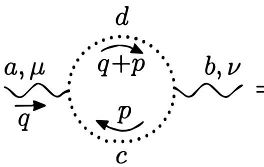
$$= -g^2 C_2(G)\delta^{ab}\int_0^1 dx \int \frac{d^4P}{(2\pi)^4}\frac{1}{(P^2-\Delta)^2}g^{\mu\nu}(d-1)[P^2+(1-x)^2q^2].$$

We can now Wick-rotate and integrate over P to obtain



$$\begin{aligned} &= \frac{ig^2}{(4\pi)^{d/2}}C_2(G)\delta^{ab}\int_0^1 dx\,\frac{1}{\Delta^{2-d/2}} \\ &\quad\times\Big(-\Gamma(1-\tfrac{d}{2})\,g^{\mu\nu}q^2\,[\tfrac{1}{2}d(d-1)x(1-x)] \\ &\qquad -\Gamma(2-\tfrac{d}{2})\,g^{\mu\nu}q^2\,[(d-1)(1-x)^2]\Big). \end{aligned} \tag{16.65}$$

Expressions (16.62) and (16.65), by themselves, do not add to any reasonable value: The pole at $d=2$ does not cancel, and the sum does not have a transverse Lorentz structure. To bring the gauge boson self-energy into its desired form, we must include the diagram with a ghost loop. According to the rules shown in Fig. 16.5, this diagram is



$$= (-1)\int\frac{d^4p}{(2\pi)^4}\frac{i}{p^2}\frac{i}{(p+q)^2}\,g^2 f^{dac}(p+q)^\mu f^{cbd}p^\nu. \tag{16.66}$$

There is no symmetry factor in this case, but there is a factor of -1 because the ghost fields anticommute. The ghost diagram can be simplified using the same set of tricks that we applied to the previous two: combine denominators, shift the integral to P, Wick-rotate, and integrate over P using dimensional regularization. The result is



$$\begin{aligned} &= \frac{ig^2}{(4\pi)^{d/2}}C_2(G)\delta^{ab}\int_0^1 dx\,\frac{1}{\Delta^{2-d/2}} \\ &\quad\times\Big(-\Gamma(1-\tfrac{d}{2})\,g^{\mu\nu}q^2\,[\tfrac{1}{2}x(1-x)] \\ &\qquad +\Gamma(2-\tfrac{d}{2})\,q^\mu q^\nu\,[x(1-x)]\Big). \end{aligned} \tag{16.67}$$

Now we are ready to put these results together. In the sum of the three diagrams, the coefficient of $\Gamma(1-\frac{d}{2})g^{\mu\nu}q^2x(1-x)$ is

$$\tfrac{1}{2}(3d-3-d^2+d-1) = (1-\tfrac{d}{2})(d-2). \tag{16.68}$$

The first factor cancels the pole of the gamma function at $d=2$. Thus, the sum of the three diagrams has no quadratic divergence and no gauge boson mass renormalization. Notice that the ghost diagram plays an essential role in this cancellation.

After the pole at $d = 2$ is canceled, $\Gamma(1-\frac{d}{2})$ becomes $\Gamma(2-\frac{d}{2})$. This term therefore combines with the others that are proportional to $\Gamma(2-\frac{d}{2})g^{\mu\nu}q^2$, to give a total coefficient of

$$(d-2)x(1-x) + \tfrac{1}{2}(2-x)^2 + \tfrac{1}{2}(1+x)^2 - (d-1)(1-x)^2. \tag{16.69}$$

Since the best way to simplify this expression is not obvious, let us put it aside and work first with the coefficient of $\Gamma(2-\frac{d}{2})q^\mu q^\nu$:

$$-(1-\tfrac{d}{2})(1-2x)^2 - (1+x)(2-x) + x(1-x) = -(1-\tfrac{d}{2})(1-2x)^2 - 2.$$

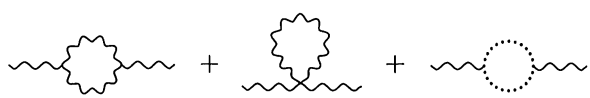
If the total self-energy is to be proportional to $(g^{\mu\nu}q^2 - q^\mu q^\nu)$, it must be possible to reduce expression (16.69) to this same form (times -1). To do so, note that Δ is symmetric with respect to $x \leftrightarrow (1-x)$, and therefore we can substitute $(1-x)$ for x in any term of the numerator. In particular, terms that are linear in x can be transformed as follows:

$$x \longrightarrow \tfrac{1}{2}x + \tfrac{1}{2}(1-x) = \tfrac{1}{2}.$$

In the end, the sum of the three pure-gauge diagrams simplifies to

$$\frac{ig^2}{(4\pi)^{d/2}}C_2(G)\delta^{ab}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}(g^{\mu\nu}q^2 - q^\mu q^\nu)\big[(1-\tfrac{d}{2})(1-2x)^2 + 2\big]. \tag{16.70}$$

This expression is manifestly transverse, as required by the Ward identity of the non-Abelian gauge theory. For future reference, we record the ultraviolet divergent part of (16.70):



$$= i(q^2g^{\mu\nu} - q^\mu q^\nu)\delta^{ab}\Big(\frac{-g^2}{(4\pi)^2}\cdot\Big(-\frac{5}{3}\Big)C_2(G)\Gamma(2-\tfrac{d}{2}) + \cdots\Big). \tag{16.71}$$

As we noted above, the result (16.70) depends on the gauge used in the calculation. In any gauge, the boson self-energy is transverse and free of quadratic divergences. However, the coefficient of the transverse Lorentz structure may depend on ξ. It turns out that, for a general value of ξ, the coefficient of the ultraviolet divergence in (16.71) is modified according to

$$-\frac{5}{3} \longrightarrow -\Big(\frac{13}{6} - \frac{\xi}{2}\Big). \tag{16.72}$$

The fact that the boson self-energy depends on the gauge does not contradict the general theorem that S-matrix elements are independent of ξ. The full set of one-loop corrections to a gauge theory S-matrix element always involves a number of different radiative corrections to vertices and propagators; the gauge dependence cancels in an intricate fashion among these various terms.

The β Function

The simplest calculation that involves a gauge-invariant combination of radiative corrections is the computation of the leading term of the Callan-Symanzik β function of a non-Abelian gauge theory. The invariance of the leading term of β could be argued intuitively, by saying that the coupling constant of the gauge theory should not evolve to large values in one scheme of calculation while it stays small in another scheme. In Section 17.2 we will demonstrate this result more cleanly by showing that the leading coefficient of the β function can be extracted from a physical cross section and so must be gauge independent. (Surprisingly, this conclusion actually applies to the first *two* coefficients of the β function, written as a power series in g.)

Recall from Section 12.2 that the β function gives the rate at which the renormalized coupling constant changes as the renormalization scale M is increased. Since Green's functions depend on M through the counterterms that subtract ultraviolet divergences, β can be computed from the counterterms that enter an appropriately chosen Green's function. For example, in Eq. (12.58), we saw that the β function of QED can be computed from the counterterms for the electron-photon vertex, the electron self-energy, and the photon self-energy. The same derivation goes through in the case of a non-Abelian gauge theory. Thus, to lowest order,

$$\beta(g) = gM\frac{\partial}{\partial M}\left(-\delta_1 + \delta_2 + \tfrac{1}{2}\delta_3\right), \tag{16.73}$$

with the conventions for the counterterm vertices shown in Fig. 16.8. In QED, the first two terms cancel by the Ward identity, so β depends only on δ_3. In the non-Abelian case, all three terms contribute. The most difficult to compute is δ_3, but we have nearly done so already by computing the gauge-boson self-energy diagrams. Let us now complete this calculation of the β function of non-Abelian gauge theory.

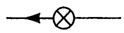
In order for the counterterm δ_3 to cancel the divergence of Eqs. (16.59) and (16.71), it must be of the form

$$\delta_3 = \frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}\left[\frac{5}{3}C_2(G) - \frac{4}{3}n_f C(r)\right], \tag{16.74}$$

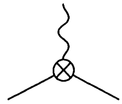
where M is the renormalization scale. Depending on the precise renormalization conditions used, there may be additional finite contributions to δ_3, but these do not contribute to the β function (to one-loop order). Similarly, the finite parts of δ_2 and δ_1 will depend on the details of the renormalization scheme. However, as we saw in Section 12.2, the one-loop contribution to the β function is the same in any scheme in which amplitudes are renormalized at a point where all momentum invariants are of the same order M^2. In dimensional regularization, a logarithmic divergence always takes the form $\Gamma(2-\frac{d}{2})/\Delta^{2-d/2}$, where Δ is some combination of momentum invariants. Thus, to compute the β function, we can simply set $\Delta = M^2$ in such expressions.



$$= -i\left(k^2 g^{\mu\nu} - k^\mu k^\nu\right)\delta^{ab}\delta_3$$



$$= i\not{p}\,\delta_2$$



$$= igt^a\gamma^\mu\delta_1$$

Figure 16.8. Counterterms needed for computing fermion interactions in a non-Abelian gauge theory.

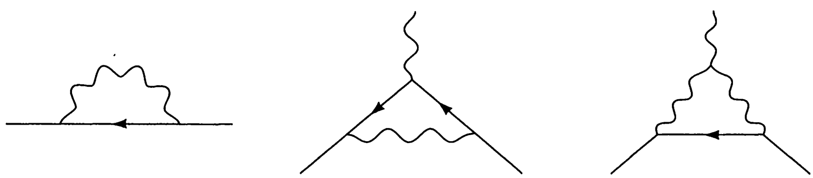
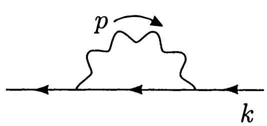


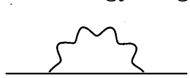
Figure 16.9. Diagrams whose divergences are subtracted by the counterterms δ_2 and δ_1.

To complete the computation of the β function, we must compute δ_2 and δ_1 to the same level of approximation. The fermion self-energy counterterm δ_2 cancels the divergence proportional to $\not{k}$ in the first diagram of Fig. 16.9. In Feynman-'t Hooft gauge, the value of this diagram is



$$= \int \frac{d^4p}{(2\pi)^4}(ig)^2\gamma^\mu t^a \frac{i(\not{p}+\not{k})}{(p+k)^2}\gamma_\mu t^a \frac{-i}{p^2}. \tag{16.75}$$

Since the divergence in the field strength renormalization is independent of the fermion mass, we have simplified (16.75) by setting the mass to zero. The product of group matrices equals the quadratic Casimir operator, by definition (15.92). The Dirac matrix structure can be reduced using a contraction identity (7.89). The rest of the calculation follows the same steps as for the boson self-energy diagrams:



$$\begin{aligned}
&= g^2C_2(r)(d-2)\int\frac{d^4p}{(2\pi)^4}\frac{(\not{p}+\not{k})}{(p+k)^2p^2}\\
&= g^2C_2(r)(d-2)\int_0^1 dx\int\frac{d^4P}{(2\pi)^4}\frac{(1-x)\not{k}}{(P^2-\Delta)^2}\\
&= \frac{ig^2}{(4\pi)^{d/2}}C_2(r)\not{k}\int_0^1 dx\,(1-x)(d-2)\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}
\end{aligned}$$

$$= \frac{ig^2}{(4\pi)^2} \not{k} C_2(r) \Gamma(2-\tfrac{d}{2}) + \cdots. \tag{16.76}$$

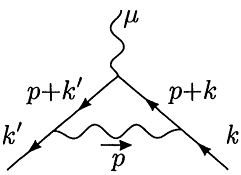
(Here $P = p + xk$ and $\Delta = -x(1-x)k^2$.)

The divergent part of this expression must be canceled by the second counterterm diagram of Fig. 16.8. Thus, if the renormalization scale is M, the counterterm must be

$$\delta_2 = -\frac{g^2}{(4\pi)^2} \frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}} \cdot C_2(r), \tag{16.77}$$

plus finite terms. We note that, like δ_3, δ_2 depends on the gauge; for example, δ_2 has no one-loop divergence in Landau gauge ($\xi = 0$).

To determine δ_1, we must compute the second and third diagrams of Fig. 16.9. The second diagram, computed in Feynman-'t Hooft gauge and for massless fermions, is



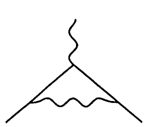
$$= \int \frac{d^4p}{(2\pi)^4}\, g^3\, t^b t^a t^b\, \frac{\gamma^\nu(\not{p} + \not{k}')\gamma^\mu(\not{p} + \not{k})\gamma_\nu}{(p + k')^2(p + k)^2 p^2}. \tag{16.78}$$

The gauge group matrices can be simplified according to

$$\begin{aligned} t^b t^a t^b &= t^b t^b t^a + t^b [t^a, t^b] \\ &= C_2(r) t^a + i t^b f^{abc} t^c \\ &= C_2(r) t^a + \tfrac{1}{2} i f^{abc} \cdot i f^{bcd} t^d \\ &= [C_2(r) - \tfrac{1}{2} C_2(G)] t^a. \end{aligned} \tag{16.79}$$

In the third line we have used the antisymmetry of f^{abc} to rewrite the matrix product as a commutator; in the last line we have used Eq. (15.93).

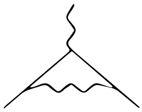
The diagrams computed earlier in this section had positive superficial degrees of divergence, so we needed to extract their logarithmic divergences carefully. The integral in (16.78), however, is superficially logarithmically divergent, and so the coefficient of this divergence can be extracted easily by considering the limit in which the integration variable p is much greater than any external momentum. In this limit, the diagram is estimated as follows:



$$\sim g^3 [C_2(r) - \tfrac{1}{2} C_2(G)] t^a \int \frac{d^4p}{(2\pi)^4} \frac{\gamma^\nu \not{p} \gamma^\mu \not{p} \gamma_\nu}{p^2 \cdot p^2 \cdot p^2}. \tag{16.80}$$

If we replace $p^\rho p^\sigma$ by $g^{\rho\sigma} p^2/d$ in the numerator of (16.80), this expression

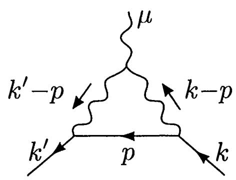
simplifies easily:



$$\sim g^3[C_2(r) - \tfrac{1}{2}C_2(G)]t^a(2-d)^2\frac{1}{d}\gamma^\mu \int \frac{d^4p}{(2\pi)^4}\frac{1}{(p^2)^2}$$
$$\sim \frac{ig^3}{(4\pi)^2}[C_2(r) - \tfrac{1}{2}C_2(G)]t^a\gamma^\mu\big(\Gamma(2-\tfrac{d}{2}) + \cdots\big). \tag{16.81}$$

This estimate gives the correct coefficient of the divergent term. It drops completely the finite terms in the vertex function, but we do not need these to compute the β function.

The third diagram of Fig. 16.9 can be analyzed in the same way. Its value, in Feynman-'t Hooft gauge and for massless fermions, is

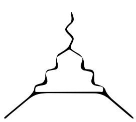


$$= \int \frac{d^4p}{(2\pi)^4}(ig\gamma_\nu t^b)\frac{i\not{p}}{p^2}(ig\gamma_\rho t^c)\frac{-i}{(k'-p)^2}\frac{-i}{(k-p)^2}$$
$$\times\, gf^{abc}\big[g^{\mu\nu}(2k'-k-p)^\rho + g^{\nu\rho}(-k'-k+2p)^\mu$$
$$+\, g^{\rho\mu}(2k-k'-p)^\nu\big]. \tag{16.82}$$

The gauge matrix product can be reduced as follows:

$$f^{abc}t^bt^c = \frac{1}{2}f^{abc}\cdot if^{bcd}t^d = \frac{i}{2}C_2(G)t^a.$$

Again we can determine the logarithmic divergence of this diagram by neglecting all external momenta in comparison with p. A straightforward calculation then yields



$$\sim \frac{g^3}{2}C_2(G)t^a\int\frac{d^4p}{(2\pi)^4}\gamma_\nu\not{p}\gamma_\rho\frac{g^{\mu\nu}p^\rho - 2g^{\nu\rho}p^\mu + g^{\rho\mu}p^\nu}{(p^2)^3}$$
$$\sim \frac{g^3}{2}C_2(G)t^a\frac{1}{d}\int\frac{d^4p}{(2\pi)^4}\frac{1}{(p^2)^2}\big[\gamma^\mu\gamma^\rho\gamma_\rho - 2\gamma^\rho\gamma^\mu\gamma_\rho + \gamma^\sigma\gamma_\sigma\gamma^\mu\big]$$
$$\sim \frac{ig^3}{(4\pi)^2}\frac{3}{2}C_2(G)\,t^a\gamma^\mu\big(\Gamma(2-\tfrac{d}{2}) + \cdots\big). \tag{16.83}$$

In the second line we have again replaced $p^\rho p^\sigma$ with $g^{\rho\sigma}p^2/d$.

The sum of the divergences in results (16.81) and (16.83) must be canceled by the third counterterm diagram in Fig 16.8. With a renormalization scale of M, we find

$$\delta_1 = -\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}\big[C_2(r) + C_2(G)\big]. \tag{16.84}$$

Notice that δ_1 is not equal to δ_2, as would have been true in the Abelian case; here δ_1 has an extra term, proportional to $C_2(G)$.

We are now ready to compute the β function. Plugging the three counterterms (16.74), (16.77), and (16.84) into our formula (16.73), we find

$$\beta(g) = (-2)\frac{g^3}{(4\pi)^2}\Big[\big(C_2(r)+C_2(G)\big) - C_2(r) + \frac{1}{2}\Big(\frac{5}{3}C_2(G) - \frac{4}{3}n_f C(r)\Big)\Big];$$

that is,

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\Big[\frac{11}{3}C_2(G) - \frac{4}{3}n_f C(r)\Big]. \tag{16.85}$$

Notice that, at least for small values of n_f, the β function is *negative* and so non-Abelian gauge theories are asymptotically free. This is a result of exceptional physical importance, first discovered by 't Hooft, Politzer, and Gross and Wilczek.* We will discuss the physical interpretation of this result further in Section 16.7, and in the next several chapters. However, for the rest of this section, we will resist the temptation to pursue the physics and instead complete our technical analysis of the divergences of non-Abelian gauge theories.

Relations among Counterterms

In the analysis just completed, we computed the β function of a non-Abelian gauge theory from the divergences of the fermion vertex and field strength renormalizations. One might visualize that we were computing the running of the coupling constant at the fermion-gauge boson vertex. Alternatively, we could have studied the divergences of the three-gauge-boson vertex or the four-gauge-boson vertex, and thus computed the running of these coupling constants. However, we saw already in Section 16.1 that non-Abelian gauge invariance knits together these separate coupling constants and requires their equality. Thus we might expect that these different calculations should produce the same value of the β function.

To clarify this issue, let us carefully enumerate all the counterterms that appear in a non-Abelian gauge theory. We start from the Lagrangian (16.34), regarded as a combination of bare fields and a bare coupling constant. In the following discussion, we denote bare quantities by the subscript 0. Then,

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}(\partial_\mu A^a_{0\nu} - \partial_\nu A^a_{0\mu})^2 + \bar\psi_0(i\not\partial - m_0)\psi_0 - \bar c^a_0\partial^2 c^a_0\\ &+ g_0 A^a_{0\mu}\bar\psi_0\gamma^\mu t^a\psi_0 - g_0 f^{abc}(\partial_\mu A^a_{0\nu})A^{b\mu}_0 A^{c\nu}_0\\ &- \frac{1}{4}g_0^2(f^{eab}A^a_{0\mu}A^b_{0\nu})(f^{ecd}A^{c\mu}_0 A^{d\nu}_0) - g_0\bar c^a_0 f^{abc}\partial^\mu A^b_{0\mu}c^c_0.\end{aligned} \tag{16.86}$$

We choose $\xi = \infty$ for simplicity. We now rescale the fields to the renormalized field strengths by extracting the factors Z_2, Z_3, Z^c_2 for the fermions, gauge

*G. 't Hooft, unpublished; H. D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973); D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973).

bosons, and ghosts, and shift the coupling to the renormalized coupling g. The Lagrangian then takes the form

$$\mathcal{L} = \mathcal{L}_{\text{ren}} + \mathcal{L}_{\text{c.t.}},$$

where $\mathcal{L}_{\text{ren}}$ is the Lagrangian (16.34) and $\mathcal{L}_{\text{c.t.}}$ takes the form

$$\begin{aligned}\mathcal{L}_{\text{c.t.}} = &-\frac{1}{4}\delta_3(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)^2 + \bar{\psi}(i\delta_2 \not{\partial} - \delta_m)\psi - \delta^c_2 \bar{c}^a \partial^2 c^a \\ &+ g\delta_1 A^a_\mu \bar{\psi}\gamma^\mu \psi - g\delta^{3g}_1 f^{abc}(\partial_\mu A^a_\nu) A^b_\mu A^c_\nu \\ &- \frac{1}{4} g^2 \delta^{4g}_1 (f^{eab} A^a_\mu A^b_\nu)(f^{ecd} A^c_\mu A^d_\nu) - g\delta^c_1 \bar{c}^a f^{abc} \partial^\mu A^b_\mu c^c,\end{aligned} \tag{16.87}$$

with the counterterms defined by

$$\begin{gathered}\delta_2 = Z_2 - 1, \qquad \delta_3 = Z_3 - 1, \qquad \delta^c_2 = Z^c_2 - 1, \qquad \delta_m = Z_2 m_0 - m, \\ \delta_1 = \frac{g_0}{g} Z_2 (Z_3)^{1/2} - 1, \qquad \delta^{3g}_1 = \frac{g_0}{g}(Z_3)^{3/2} - 1, \\ \delta^{4g}_1 = \frac{g_0^2}{g^2}(Z_3)^2 - 1, \qquad \delta^c_1 = \frac{g_0}{g} Z^c_2 (Z_3)^{1/2} - 1.\end{gathered} \tag{16.88}$$

Notice that these eight counterterms depend on five underlying parameters; thus, there are three relations among them. The situation is very similar to that for the scalar theories with spontaneously broken symmetry that we studied in Chapter 11. The underlying symmetry of the theory—here, local gauge invariance—implies relations among the divergent amplitudes of the theory and among the counterterms required to cancel them. In the present case, a set of five renormalization conditions uniquely specifies all of the counterterms in a way that removes all divergences from the theory.

This program is especially simple at one-loop order. In this case we can expand g_0/g and the various Z factors about 1, keeping only the leading-order contribution to each counterterm. Then the three relations among the counterterms can be written

$$\delta_1 - \delta_2 = \delta^{3g}_1 - \delta_3 = \tfrac{1}{2}(\delta^{4g}_1 - \delta_3) = \delta^c_1 - \delta^c_2. \tag{16.89}$$

It is instructive to check explicitly that the values of δ^{3g}_1, δ^{4g}_1, and δ^c_1 determined from (16.89) indeed remove the divergences of the corresponding vertex diagrams; this is the subject of Problem 16.3. Using relations (16.89), it is easy to show that the one-loop calculation of the β function will yield the same value, whichever gauge boson vertex is used in the computation. More generally, consider a non-Abelian gauge theory with many different species of particles, bosons and fermions, which couple to the gauge field. Then, to one-loop order, the quantity

$$\delta^i_1 - \delta^i_2,$$

where δ_1^i is the vertex counterterm for species i and δ_2^i is the corresponding field strength counterterm, takes a universal value. This value is gauge dependent, so that the gauge dependence of its divergent part cancels the gauge dependence of δ_3 in the computation of the β function.

In our discussion of the counterterms of QED at the end of Section 10.3, we remarked that the relation between δ_1 and δ_2 insured that all electrically charged species see a common universal value of the coupling constant e. In non-Abelian gauge theories, the relations (16.89) and their higher-loop generalizations preserve the universality of the non-Abelian couplings. In QED, we were able to obtain an even stronger relation, $\delta_1 = \delta_2$ or $Z_1 = Z_2$, from the absolute normalization of the matrix elements of the vector current. However, in non-Abelian gauge theories, the corresponding vector current $j^{\mu a} = \bar{\psi}\gamma^\mu t^a \psi$ transforms under local gauge transformations in the adjoint representation. Thus the Faddeev-Popov prescription cannot be used to compute matrix elements of this current unambiguously, and thus the normalization of these matrix elements is not preserved by the perturbation theory.

16.6 Asymptotic Freedom: The Background Field Method

In the previous section, we saw that the β function of a non-Abelian gauge theory with a sufficiently small number of fermions is negative. This result is important enough that it is worthwhile to derive it twice. The preceding derivation was straightforward but not very illuminating. In this section we give a second derivation of the same result, which is more abstract but much cleaner and more transparent.

The method of this section reflects the spirit of Wilson's idea of integrating out the high-momentum degrees of freedom, while taking proper care to preserve gauge invariance. We will compute the effective action of a non-Abelian gauge theory for a fixed, slowly varying, classical background gauge field $A_\mu^a(x)$. By adopting a canonical normalization of this field, we can interpret the coefficient of the effective action as a running coupling constant. This method is analogous to Polyakov's method for computing the β function of the nonlinear sigma model, presented in Section 13.3.

Background Field Perturbation Theory

To set up the computation, rescale the gauge field $gA_\mu^a \to A_\mu^a$. In this normalization, the gauge coupling is removed from the covariant derivative and moved to the coefficient of the gauge field kinetic energy term. We thus start from the Lagrangian

$$\mathcal{L} = -\frac{1}{4g^2}(F_{\mu\nu}^a)^2 + \bar{\psi}(i\not{D})\psi, \tag{16.90}$$

with

$$\begin{aligned} D_\mu &= \partial_\mu - iA^a_\mu t^a, \\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu, \end{aligned} \tag{16.91}$$

and the fermion mass set to zero for simplicity. The transformation laws of A^a_μ and ψ are also independent of the coupling constant:

$$\delta A^a_\mu = \partial_\mu \alpha^a + f^{abc}A^b_\mu \alpha^c, \qquad \delta\psi = i\alpha^a t^a \psi. \tag{16.92}$$

On the other hand, the coupling constant g will appear in the gauge field propagator.

Next, split the gauge field into a classical background field and a fluctuating quantum field:

$$A^a_\mu \to A^a_\mu + \mathcal{A}^a_\mu. \tag{16.93}$$

We will treat the classical part A^a_μ as a fixed field configuration and the fluctuating part $\mathcal{A}^a_\mu$ as the integration variable of the functional integral. From here on, we will use the symbol D_μ to denote the covariant derivative with respect to the background field: $D_\mu = \partial_\mu - iA^a_\mu t^a$. Then

$$\bar\psi(i\not{D})\psi \to \bar\psi(i\not{D})\psi + \mathcal{A}^a_\mu \bar\psi\gamma^\mu t^a \psi. \tag{16.94}$$

The Yang-Mills field strength decomposes as follows:

$$\begin{aligned} F^a_{\mu\nu} \to\ & \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu \\ & + \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + f^{abc}(A^b_\mu \mathcal{A}^c_\nu - A^b_\nu \mathcal{A}^c_\mu) + f^{abc}\mathcal{A}^b_\mu \mathcal{A}^c_\nu \\ = & F^a_{\mu\nu} + D_\mu \mathcal{A}^a_\nu - D_\nu \mathcal{A}^a_\mu + f^{abc}\mathcal{A}^b_\mu \mathcal{A}^c_\nu, \end{aligned} \tag{16.95}$$

where, in the last line, $F^a_{\mu\nu}$ is the field strength of the classical field, and D_μ is the covariant derivative in the adjoint representation, Eq. (15.86). Notice that, both in (16.94) and in (16.95), the derivative ∂_μ appears only as a part of the covariant derivative with respect to the background field.

If the background field A^a_μ is regarded as fixed, the Lagrangian has a local gauge symmetry implemented by transformations on $\mathcal{A}^a_\mu$:

$$\mathcal{A}^a_\mu \to \mathcal{A}^a_\mu + D_\mu \alpha^a + f^{abc}\mathcal{A}^b_\mu \alpha^c. \tag{16.96}$$

To define the functional integral, we must gauge-fix using the Faddeev-Popov procedure. We choose a gauge-fixing condition that is covariant with respect to the background gauge field:

$$G^a(A) = D^\mu \mathcal{A}^a_\mu - \omega^a, \tag{16.97}$$

instead of (16.28). The Faddeev-Popov determinant involves the variation of this operator with respect to the gauge transformation (16.96). As in Section 16.2, we can promote the gauge-fixing term to the exponent, to quantize the

theory in the background field analogue of Feynman-'t Hooft gauge. Then the gauge-fixed Lagrangian is

$$\begin{aligned}\mathcal{L}_{\text{FP}} = &-\frac{1}{4g^2}\left(F^a_{\mu\nu} + D_\mu\mathcal{A}^a_\nu - D_\nu\mathcal{A}^a_\mu * f^{abc}\mathcal{A}^b_\mu\mathcal{A}^c_\nu\right)^2 - \frac{1}{2g^2}(D^\mu\mathcal{A}^a_\mu)^2 \\ &+ \bar{\psi}(i\not{D} + \mathcal{A}^a_\mu\gamma^\mu t^a)\psi - \bar{c}^a(D^2)^{ac}c^c - \bar{c}^a D^\mu(f^{abc}\mathcal{A}^b_\mu c^c).\end{aligned} \tag{16.98}$$

The Lagrangian (16.98) is gauge-fixed, but it is invariant under a local symmetry that transforms both $\mathcal{A}^a_\mu$ and the background field A^a_μ:

$$\begin{aligned}A^a_\mu &\rightarrow A^a_\mu + D_\mu\beta^a \\ \mathcal{A}^a_\mu &\rightarrow \mathcal{A}^a_\mu - f^{abc}\beta^b\mathcal{A}^c_\mu \\ \psi &\rightarrow \psi + i\beta^a t^a\psi \\ c^a &\rightarrow c^a - f^{abc}\beta^b c^c.\end{aligned} \tag{16.99}$$

Under this transformation, $\mathcal{A}^a_\mu$ transforms as a matter field in the adjoint representation, while A^a_μ carries the part of the local gauge transformation proportional to $\partial_\mu\beta^a$. To prove that (16.99) is a symmetry of (16.98), we need only note that (16.98) is globally invariant, and that A^a_μ appears in (16.98) only as a part of the covariant derivative and the field strength. The transformation (16.99) is also a symmetry of the functional measure. Thus, if we functionally integrate over $\mathcal{A}^a_\mu$, ψ, and c^a to compute the effective action, the result must be invariant to local gauge transformations of A^a_μ. This observation greatly simplifies the analysis of the effective action.

One-Loop Correction to the Effective Action

Let us now compute the effective action, using the method of Section 11.4. To compute $\Gamma[A^a_\mu]$ to one-loop order, we drop terms linear in the fluctuating field $\mathcal{A}^a_\mu$ and then integrate over the terms quadratic in $\mathcal{A}^a_\mu$ and the fermion and ghost fields. This produces functional determinants, which we can evaluate into an appropriate form for an effective action.

To carry out this program, we must work out the terms in (16.98) quadratic in each of the various fields. The terms quadratic in $\mathcal{A}^a_\mu$ are:

$$\mathcal{L}_\mathcal{A} = -\frac{1}{2g^2}\left\{\tfrac{1}{2}(D_\mu\mathcal{A}^a_\nu - D_\nu\mathcal{A}^a_\mu)^2 + F^{a\mu\nu}f^{abc}\mathcal{A}^b_\mu\mathcal{A}^c_\nu + (D^\mu\mathcal{A}^a_\mu)^2\right\}. \tag{16.100}$$

After integrating by parts, we can rewrite this as

$$\mathcal{L}_\mathcal{A} = -\frac{1}{2g^2}\left\{\mathcal{A}^a_\mu[-(D^2)^{ab}g^{\mu\nu} + (D^\nu D^\mu)^{ab} - (D^\mu D^\nu)^{ab}]\mathcal{A}^b_\nu - \mathcal{A}^a_\mu f^{abc}F^{b\mu\nu}\mathcal{A}^c_\nu\right\}. \tag{16.101}$$

The term in brackets contains the commutator of covariant derivatives. This can be simplified using (15.48); the result combines with the last term to give

$$\mathcal{L}_\mathcal{A} = -\frac{1}{2g^2}\left\{\mathcal{A}^a_\mu[-(D^2)^{ac}g^{\mu\nu} - 2f^{abc}F^{b\mu\nu}]\mathcal{A}^c_\nu\right\}. \tag{16.102}$$

The first term is part of a covariant d'Alembertian operator. The second term seems quite special, but we can put it into a form that will be convenient later as follows: First, we recognize that $F^b_{\mu\nu}$ is contracted with a group generator in the adjoint representation. Next, we introduce the matrix (3.18) that is the generator of Lorentz transformations on 4-vectors:

$$(\mathcal{J}^{\rho\sigma})_{\alpha\beta} = i(\delta^\rho_\alpha\delta^\sigma_\beta - \delta^\sigma_\alpha\delta^\rho_\beta). \tag{16.103}$$

With these replacements, we can write (16.102) in the form

$$\mathcal{L}_\mathcal{A} = -\frac{1}{2g^2}\{\mathcal{A}^a_\mu[-(D^2)^{ac}g^{\mu\nu} + 2(\tfrac{1}{2}F^b_{\rho\sigma}\mathcal{J}^{\rho\sigma})^{\mu\nu}(t^b_G)^{ac}]\mathcal{A}^c_\nu\}. \tag{16.104}$$

The object in brackets can be considered as a generalized d'Alembertian for fluctuations on the background field.

Next, we reduce the quadratic terms in fermion fields in a similar way. The quadratic Lagrangian for the fermion field is

$$\mathcal{L}_\psi = \bar{\psi}(i\not{D})\psi. \tag{16.105}$$

Integrating over the fermion fields, we find the determinant of the operator $(i\not{D})$. This is conveniently expressed as the square root of the determinant of the operator

$$\begin{aligned}(i\not{D})^2 &= -\gamma^\mu\gamma^\nu D_\mu D_\nu \\ &= \left(-\tfrac{1}{2}\{\gamma^\mu,\gamma^\nu\} - \tfrac{1}{2}[\gamma^\mu,\gamma^\nu]\right)D_\mu D_\nu \\ &= -D^2 + 2i\left(\frac{i}{4}[\gamma^\mu,\gamma^\nu]\right)D_\mu D_\nu.\end{aligned} \tag{16.106}$$

In the last line, the commutator of Dirac matrices forms the generator of Lorentz transformations in the spinor representation, $S^{\mu\nu}$ (3.23). Since this object is antisymmetric in its indices, the product $D_\mu D_\nu$ that is contracted with it can be replaced by half of their commutator. Then (16.106) takes the form

$$(i\not{D})^2 = -D^2 + 2(\tfrac{1}{2}F^b_{\rho\sigma}S^{\rho\sigma})t^b, \tag{16.107}$$

where t^a is now given in the representation of the fermions. This is just the d'Alembertian in (16.104), rewritten for the new set of spin and gauge quantum numbers. If the theory contains n_f species of fermions, the fermionic functional integral gives the determinant of (16.107) raised to the power $n_f/2$.

The quadratic term in ghosts is simply

$$\mathcal{L}_c = \bar{c}^a[-(D^2)^{ab}]c^b; \tag{16.108}$$

This contains the same d'Alembertian operator written for the case of spin zero.

To summarize all of these results, we define the general covariant background-field d'Alembertian as

$$\Delta_{r,j} = -D^2 + 2(\tfrac{1}{2}F^b_{\rho\sigma}\mathcal{J}^{\rho\sigma})t^b, \tag{16.109}$$

acting on a field of representation r and spin j. The square of the covariant derivative gives the normal, convective, minimal coupling of the particle described by $\Delta_{r,j}$ to the gauge field. The additional term is a magnetic moment interaction with the gauge field, whose strength corresponds to a g-factor $g = 2$. Using this general expression, we can write the effective action for the classical fields A^a_μ, to one-loop order, as

$$\begin{aligned} e^{i\Gamma[A]} &= \int \mathcal{D}\mathcal{A}\mathcal{D}\psi\mathcal{D}c \exp\Big[i\int d^4x\,(\mathcal{L}_{\text{FP}} + \mathcal{L}_{\text{c.t.}})\Big] \\ &= \exp\Big[i\int d^4x\big(-\frac{1}{4g^2}(F^a_{\mu\nu})^2 + \mathcal{L}_{\text{c.t.}}\big)\Big] \\ &\qquad \cdot \big(\det\Delta_{G,1}\big)^{-1/2}\big(\det\Delta_{r,1/2}\big)^{+n_f/2}\,\big(\det\Delta_{G,0}\big)^{+1}, \end{aligned} \tag{16.110}$$

where $\mathcal{L}_{\text{c.t.}}$ is the counterterm Lagrangian and the three determinants are the results of evaluating the gauge field, fermion, and ghost functional integrals. Additional loop corrections to the effective action are suppressed by another factor of g^2.

Since each integral contributing to (16.110) is invariant to (16.99), each determinant will be a gauge-invariant functional of A^a_μ. If we expand the determinants in powers of the background field, we should then find a series of terms that begins

$$\log\det\Delta_{r,j} = i\int d^4x\,\Big(\frac{1}{4}\mathbf{C}_{r,j}(F^a_{\mu\nu})^2 + \cdots\Big), \tag{16.111}$$

where the succeeding terms contain higher-dimension gauge-invariant operators. The coefficient $\mathbf{C}_{r,j}$ can depend on the representation r and the spin j. This first term of the expansion modifies the zeroth-order effective action according to

$$\frac{1}{4g^2}(F^a_{\mu\nu})^2 \to \frac{1}{4}\Big(\frac{1}{g^2} + \frac{1}{2}\mathbf{C}_{G,1} - \mathbf{C}_{G,0} - \frac{n_f}{2}\mathbf{C}_{r,1/2}\Big)(F^a_{\mu\nu})^2. \tag{16.112}$$

The factors $\mathbf{C}_{r,j}$ are dimensionless but, since they arise from a one-loop computation, we should expect that they are logarithmically divergent:

$$\mathbf{C}_{r,j} = c_{r,j}\log\frac{\Lambda^2}{k^2} + \cdots, \tag{16.113}$$

where k is a momentum characterizing the variation of the background field. The counterterm δ_3 removes the divergence; if we impose a renormalization condition at the scale M, then the addition of (16.113) and its counterterm gives the result (16.112) with the replacement

$$\mathbf{C}_{r,j} = c_{r,j}\log\frac{M^2}{k^2} + \cdots. \tag{16.114}$$

Then the original fixed coupling constant in the effective action is replaced by a running coupling constant

$$\frac{1}{g^2(k^2)} = \frac{1}{g^2} + \left(\frac{1}{2}c_{G,1} - c_{G,0} - \frac{n_f}{2}c_{r,1/2}\right)\log\frac{M^2}{k^2}, \tag{16.115}$$

or

$$g^2(k^2) = \frac{g^2}{1 - (\frac{1}{2}c_{G,1} - c_{G,0} - \frac{n_f}{2}c_{r,1/2})g^2\log k^2/M^2}. \tag{16.116}$$

By comparing this form to Eq. (12.88), we see that this running coupling constant is the solution to the renormalization group equation for the β function

$$\beta(g) = \left(\frac{1}{2}c_{G,1} - c_{G,0} - \frac{n_f}{2}c_{r,1/2}\right)g^3. \tag{16.117}$$

Thus, by calculating the $c_{r,j}$, we can directly obtain the leading coefficient of the β function.

Computation of the Functional Determinants

To compute $c_{r,j}$, we must work out the first term in the expansion of the determinant in powers of the external field. To expand the determinant, we proceed as in the example in Section 9.5. Write

$$\Delta_{r,j} = -\partial^2 + \Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})}, \tag{16.118}$$

where

$$\begin{aligned} \Delta^{(1)} &= i[\partial^\mu A^a_\mu t^a + A^a_\mu t^a \partial^\mu] \\ \Delta^{(2)} &= A^{a\mu} t^a A^b_\mu t^b \\ \Delta^{(\mathcal{J})} &= 2(\tfrac{1}{2}F^b_{\rho\sigma}\mathcal{J}^{\rho\sigma})t^b. \end{aligned} \tag{16.119}$$

The pieces $\Delta^{(1)}$ and $\Delta^{(\mathcal{J})}$ contain one power of the external field; $\Delta^{(2)}$ contains two powers of A^a_μ. Treating these terms as perturbations, we write

$$\begin{aligned} \log\det\Delta_{r,j} &= \log\det[-\partial^2 + (\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})})] \\ &= \log\det[-\partial^2] + \log\det[1 + (-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})})] \\ &= \log\det[-\partial^2] + \operatorname{tr}\log[1 + (-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})})] \\ &= \log\det[-\partial^2] + \operatorname{tr}[(-\partial^2)^{-1}(\Delta^{(1)} + \Delta^{(2)} + \Delta^{(\mathcal{J})}) + \cdots]. \end{aligned} \tag{16.120}$$

The first term of the right in (16.120) is an irrelevant constant. The terms in this expansion that are linear in A^a_μ vanish by gauge invariance (or, more explicitly, because $\operatorname{tr}[t^a] = 0$). The quadratic terms in A^a_μ must organize themselves into the structure of (16.111), plus terms with higher derivatives.

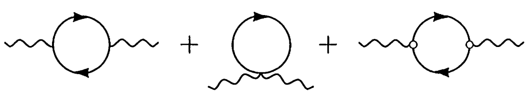


Figure 16.10. Terms quadratic in the external field in the expansion of $\log\det\Delta_{r,j}$. The special vertex arises from the $F^{\rho\sigma}\mathcal{J}_{\rho\sigma}$ coupling.

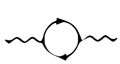
The terms in (16.111) quadratic in A^a_μ can be written in Fourier space as

$$\log\det\Delta_{r,j} = \frac{i}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu(-k)A^b_\nu(k)(k^2g^{\mu\nu}-k^\mu k^\nu)\cdot\big[\mathbf{C}_{r,j}+\mathcal{O}(k^2)\big]. \tag{16.121}$$

We will now compute these terms explicitly from (16.120) and bring them into the form of (16.121). The terms with two powers of A^a_μ in the expansion (16.120) are those with one power of $\Delta^{(2)}$ or two powers of $\Delta^{(1)}$ or $\Delta^{(\mathcal{J})}$. Further, terms linear in $\Delta^{(\mathcal{J})}$ are proportional to $\mathrm{tr}[\mathcal{J}^{\rho\sigma}]=0$, so the cross term between these two structures vanishes. The three remaining contributions correspond to the Feynman diagrams shown in Fig. 16.10.

The term involving two powers of $\Delta^{(1)}$ is

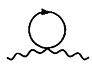
$$-\tfrac{1}{2}\,\mathrm{tr}\big[(-\partial^2)^{-1}\Delta^{(1)}(-\partial^2)^{-1}\Delta^{(1)}\big] =$$



$$= -\tfrac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\,\mathrm{tr}\,\frac{1}{p^2}(2p+k)^\mu t^a\frac{1}{(p+k)^2}(2p+k)^\nu t^b, \tag{16.122}$$

where the trace is now simply a trace over gauge and spin indices. The factor 1/2 comes from the expansion of the logarithm. The term involving one power of $\Delta^{(2)}$ is

$$\mathrm{tr}\big[(-\partial^2)^{-1}\Delta^{(2)}\big] =$$



$$= \int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\,\mathrm{tr}\,\frac{1}{p^2}g^{\mu\nu}t^at^b. \tag{16.123}$$

As Fig. 16.10 suggests, these two contributions are precisely proportional to the contribution of a scalar particle to the QED vacuum polarization, times the factor

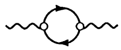
$$\mathrm{tr}[t^at^b] = C(r)d(j)\delta^{ab}, \tag{16.124}$$

where $d(j)$ is the number of spin components. The values of the diagrams can be worked out using the methods of the previous section (or simply recalled from Problem 9.1). One finds that the two diagrams together sum up to the gauge-invariant form (16.121), to give

$$\tfrac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu(-k)A^b_\nu(k)(k^2g^{\mu\nu}-k^\mu k^\nu)\cdot\Big[i\frac{C(r)d(j)}{3(4\pi)^2}\Gamma(2-\tfrac{d}{2})+\cdots\Big]. \tag{16.125}$$

The term involving two powers of $\Delta^{(\mathcal{J})}$ is

$$-\tfrac{1}{2}\operatorname{tr}\left[(-\partial^2)^{-1}\Delta^{(\mathcal{J})}(-\partial^2)^{-1}\Delta^{(\mathcal{J})}\right] =$$



$$= -\tfrac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\operatorname{tr}\frac{1}{p^2}(2ik_\rho g_{\mu\sigma}\mathcal{J}^{\rho\sigma})t^a\frac{1}{(p+k)^2}(-2ik_\alpha g_{\nu\beta}\mathcal{J}^{\alpha\beta})t^b. \tag{16.126}$$

To evaluate this, define $C(j)$ as the trace over spin indices

$$\operatorname{tr}[\mathcal{J}^{\rho\sigma}\mathcal{J}^{\alpha\beta}] = (g^{\rho\alpha}g^{\sigma\beta} - g^{\rho\beta}g^{\sigma\alpha})C(j). \tag{16.127}$$

It is straightforward to work out from the explicit expressions that

$$C(j) = \begin{cases} 0 & \text{scalars;} \\ 1 & \text{Dirac spinors;} \\ 2 & \text{4-vectors.} \end{cases} \tag{16.128}$$

Then (16.126) can be evaluated as

$$-\frac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu A^b_\nu\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^2}\frac{1}{(p+k)^2}(k^2g^{\mu\nu} - k^\mu k^\nu)4C(r)C(j)$$
$$= \frac{1}{2}\int\frac{d^4k}{(2\pi)^4}A^a_\mu(-k)A^b_\nu(k)(k^2g^{\mu\nu} - k^\mu k^\nu)\Big(-i\frac{4C(r)C(j)}{(4\pi)^2}\Gamma(2-\tfrac{d}{2}) + \cdots\Big). \tag{16.129}$$

Adding (16.125) and (16.129), we find that the coefficient $\mathbf{C}_{r,j}$ in (16.111) is given by

$$\mathbf{C}_{r,j} = \frac{1}{(4\pi)^2}\left[\tfrac{1}{3}d(j) - 4C(j)\right]C(r)\Gamma(2-\tfrac{d}{2}). \tag{16.130}$$

Thus,

$$c_{r,j} = \frac{1}{(4\pi)^2}\left[\tfrac{1}{3}d(j) - 4C(j)\right]C(r), \tag{16.131}$$

or explicitly,

$$c_{r,j} = \frac{C(r)}{(4\pi)^2}\cdot\begin{cases} +1/3 & \text{scalars;} \\ -8/3 & \text{Dirac spinors;} \\ -20/3 & \text{4-vectors.} \end{cases} \tag{16.132}$$

Notice that, whenever the magnetic moment term is nonzero, it dominates, and that its coefficient is opposite in sign from the convective term.

Inserting the values from (16.132) into (16.117), we find

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\Big(\frac{11}{3}C_2(G) - \frac{4}{3}n_f C(r)\Big). \tag{16.133}$$

We thus confirm the conclusion of the previous section, that non-Abelian gauge theories with sufficiently few fermions are asymptotically free.

16.7 Asymptotic Freedom: A Qualitative Explanation

In the previous two sections† we twice calculated the β function in non-Abelian gauge theory:

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}C_2(G) - \frac{4}{3}n_f C(r)\right). \tag{16.134}$$

Here n_f is the number of fermion species in representation r, $C(r)$ is the constant appearing in the orthogonality relation (15.78) for the representation matrices, and $C_2(G)$ is the quadratic Casimir operator of the adjoint representation of the group, defined in Eq. (15.92). In an $SU(N)$ gauge theory with fermions in the fundamental representation, this result becomes

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}N - \frac{2}{3}n_f\right). \tag{16.135}$$

The overall minus sign implies that, for sufficiently small n_f, non-Abelian gauge theories are asymptotically free. In this case the running coupling constant tends to zero at large momenta, according to Eq. (12.92):

$$g^2(k) = \frac{g^2}{1 + \frac{g^2}{(4\pi)^2}\left(\frac{11}{3}N - \frac{2}{3}n_f\right)\log(k^2/M^2)}. \tag{16.136}$$

The asymptotic freedom of non-Abelian gauge theories is a surprising conclusion. When we first encountered the running of the electromagnetic coupling in Section 7.5, we found it easy to understand the direction of the coupling constant flow: The vacuum acquires a dielectric property due to virtual electron-positron pair creation, causing the effective electric charge to decrease at large distances. In non-Abelian gauge theories, according to Eq. (16.134), the fermions still produce such an effect. Furthermore, since the non-Abelian gauge bosons are charged, they should produce an additional screening effect. But according to Eq. (16.134), the net effect of the non-Abelian gauge bosons is *opposite* in sign. Apparently there must be other, competing, effects, which overcome the effect of screening and change the sign of the β function.

The precise form of these effects depends on the gauge. They are simplest to describe in the Coulomb gauge, for which the gauge fixing condition is

$$\partial_i A^{ai} = 0. \tag{16.137}$$

We will not work out the full quantization in this gauge; rather, we will just describe its qualitative features. As in electrodynamics, the field quanta in Coulomb gauge are described in a non-Lorentz-covariant manner as transversely polarized photons. There are no timelike or longitudinal photons and

†Section 16.7 draws on the main result of 16.5 and 16.6, but does not depend on these earlier sections. However, even if you have not read Section 16.5, you may wish to skim pages 522 through 531 to get an overview of how the β function can be calculated.

no propagating ghosts. However, there is a Coulomb potential, described by the field A^{a0}, which obeys a constraint equation analogous to Gauss's law. Not surprisingly, in the non-Abelian case, Gauss's law takes the gauge-covariant form

$$D_i E^{ai} = g\rho^a, \tag{16.138}$$

where $E^{ai} = F^{a0i}$ and ρ^a is the charge density of the global symmetry current of the fermions. Recall from Eq. (15.86) that the covariant derivative acting on a field in the adjoint representation is

$$(D_\mu \phi)^a = \partial_\mu \phi^a + g f^{abc} A^b_\mu \phi^c.$$

To analyze the consequences of Eq. (16.138), we will choose an example as simple and explicit as possible. Let the gauge group be $SU(2)$, so that $a = 1, 2, 3$ and $f^{abc} = \epsilon^{abc}$. Let us compute the Coulomb potential of a point charge of magnitude $+1$ with the orientation $a = 1$. We will solve for E^{ai} using an iteration process, putting the gauge-field term of the covariant derivative on the right-hand side of the equation:

$$\partial_i E^{ai} = g\delta^{(3)}(\mathbf{x})\delta^{a1} + g\epsilon^{abc} A^{bi} E^{ci}. \tag{16.139}$$

The second term on the right shows that, in a non-Abelian gauge theory, a region containing vector potentials and electric fields that are parallel in physical space and perpendicular in the group space is a source of electric field.

The implication of Eq. (16.139) is worked out pictorially in Fig. 16.11. The leading term on the right-hand side of (16.139) implies a $1/r^2$ electric field of type $a = 1$ radiating from $\mathbf{x} = 0$. Somewhere in space, this electric field will cross with a bit of vector potential A^{ai} arising as a fluctuation of the vacuum. For definiteness, let us assume that this fluctuation has $a = 2$ and points in some diagonal direction, as shown in Fig. 16.11(a). Then the second term on the right-hand side of Eq. (16.139) is negative for $a = 3$: There is a sink of the field E^{3i} at this location, as shown in Fig. 16.11(b). These new fields are, in two locations, parallel or antiparallel to the original A^{ai} field fluctuation. Looking again at the second term of Eq. (16.139), we see that there is a source of electric field with $a = 1$ closer to the origin, and a sink of electric field with $a = 1$ farther away. This is an induced electric dipole in the vacuum, shown in Fig. 16.11(c). But look at the signs: This dipole is oriented toward the original charge, and thus serves to amplify rather than screen it! The effect of the original charge thus becomes stronger at larger distances.

The competition between this antiscreening effect and the screening due to virtual pairs of gauge bosons must be worked out quantitatively. When this is done,[‡] one finds that the antiscreening effect is 12 times larger.

In this argument, it is a set of dynamical features peculiar to the non-Abelian gauge theory that enables the coupling constant to be amplified rather

‡T. Appelquist, M. Dine, and I. Muzinich, *Phys. Lett.* **69B**, 231 (1977).

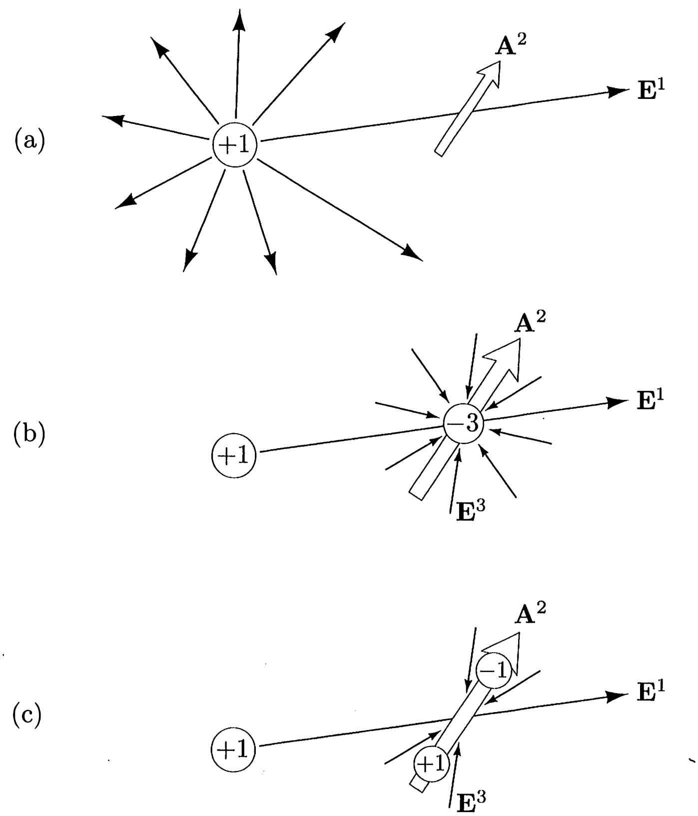


Figure 16.11. The effect of vacuum fluctuations on the Coulomb field of an $SU(2)$ gauge theory. In (a), a fluctuation $\mathbf{A}^2$ occurs on top of the $1/r^2$ field $\mathbf{E}^1$. These combined fields generate a sink of the field $\mathbf{E}^3$, as shown in (b). The $\mathbf{E}^3$ field, in turn, combines with $\mathbf{A}^2$ to create an effective $\mathbf{E}^1$ dipole, shown in (c). The dipole points toward the original charge, enhancing its field at large distances.

than screened at large distances. This suggests that asymptotic freedom might be a special property of non-Abelian gauge theories. Although the statement can be proved only by exhausting other cases, it does actually turn out to be true: Among renormalizable quantum field theories in four spacetime dimensions, only the non-Abelian gauge theories are asymptotically free.* We have already seen in Chapter 14 that asymptotic freedom was suggested experimentally as a property of the strong interactions. In the following chapter we will build a model of the strong interactions out of a non-Abelian gauge theory and explore its properties in detail.

*S. Coleman and D. J. Gross, *Phys. Rev. Lett.* **31**, 851 (1973).

Problems

16.1 Arnowitt-Fickler gauge. Perform the Faddeev-Popov quantization of Yang-Mills theory in the gauge $A^{3a} = 0$, and write the Feynman rules. Show that there are no propagating ghosts, and that the gauge field is reduced to two positive-metric degrees of freedom. (Although the gauge condition violates Lorentz invariance, this symmetry is restored in the calculation of gauge-invariant S-matrix elements.)

16.2 Scalar field with non-Abelian charge. Consider a non-Abelian gauge theory with gauge group G. Couple to this theory a complex scalar field in the representation r.

(a) Show that the Feynman rules for the scalar field are a simple modification of the Feynman rules displayed for scalar QED in Problem 9.1(a).

(b) Compute the contribution of this scalar field to the β function, and show that the full β function for this theory is

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(\frac{11}{3}C_2(G) - \frac{1}{3}C(r)\right).$$

16.3 Counterterm relations. In Section 16.5, we computed the divergent parts of δ_1, δ_2, and δ_3. It is a good exercise to compute the divergent parts of the remaining counterterms in Eq. (16.88) to one-loop order in the Feynman-'t Hooft gauge, and to explicitly verify that the counterterm relations (16.89) are consistent with the removal of ultraviolet divergences.

(a) The ghost counterterms are particularly easy to compute. Work out δ_1^c and δ_2^c, and show that the divergent part of their difference equals the divergent part of $\delta_1 - \delta_2$. This gives a derivation of asymptotic freedom that is slightly easier than the one in Section 16.5.

(b) Compute the counterterm for the 3-gauge-boson vertex and verify the first equality in (16.89).

(c) Compute the counterterm for the 4-gauge-boson vertex and find, when the smoke clears, the second relation in (16.89).

Chapter 17

Quantum Chromodynamics

The key to constructing a realistic model of the strong interactions is the phenomenon of asymptotic freedom. Chapter 14 described the experimental discovery of this phenomenon, while Chapter 16 presented the theoretical proof that non-Abelian gauge theories are asymptotically free. We are now ready to explore the consequences of these discoveries.

We will begin by arguing that the most natural candidate for a model of the strong interactions is the non-Abelian gauge theory with gauge group $SU(3)$, coupled to fermions (quarks) in the fundamental representation. This theory is known as *Quantum Chromodynamics*, or QCD. After some general discussion of QCD in Section 17.1, we will investigate a number of specific QCD scattering processes in Sections 17.2 through 17.4. The most interesting application of QCD, however, is of a somewhat more sophisticated nature; it comes in the prediction of a pattern of slow violations of the Bjorken scaling relation discussed in Chapter 14. Section 17.5 develops the additional theoretical tools that are needed to understand these violations.

Although this chapter includes many references to experiments, we remind the reader that, for QCD as for QED or critical phenomena, this book is primarily a textbook of theoretical methods rather than a review and interpretation of experimental data. The details of experimental techniques and results on strong interaction physics are reviewed in a number of excellent texts (see the bibliography). We hope that this chapter will give the theoretical foundation necessary to illuminate and interpret these results.

17.1 From Quarks to QCD

Our current theoretical picture of the strong interactions began with the identification of the elementary fermions that make up the proton and other hadrons. As the properties of these fermions became better understood, the nature of their interactions became tightly constrained, in a way that led eventually to a unique candidate theory. In order to appreciate the uniqueness of this theory, we begin this chapter with a simplified history of how it arose.

In 1963, Gell-Mann and Zweig proposed a model that explained the spectrum of strongly interacting particles in terms of elementary constituents

called *quarks*. Mesons were expected to be quark-antiquark bound states. Indeed, the lightest mesons have just the correct quantum numbers to justify this interpretation; they are spin-0 and spin-1 states of odd parity, just as we found for fermion-antifermion bound states of zero orbital angular momentum in Chapter 3. Baryons were interpreted as bound states of three quarks. To explain the electric charges and other quantum numbers of hadrons, Gell-Mann and Zweig needed to assume three species of quarks, up (u), down (d), and strange (s). Additional hadrons discovered since that time require the existence of three more species: charm (c), bottom (b), and top (t). To make baryons with integer charges, the quarks needed to be assigned fractional electric charge: $+2/3$ for u, c, t, and $-1/3$ for d, s, b. Then, for example, the proton would be a bound state of uud, while the neutron would be a bound state of udd. The six types of quarks are conventionally referred to as *flavors*.

The quark model had great success in predicting new hadronic states, and in explaining the strengths of electromagnetic and weak-interaction transitions among different hadrons. In particular, the quark model naturally incorporates the most important symmetry relations among strongly interacting particles. If one assumes that the u and d quarks have identical masses and interactions, the $SU(2)$ group that acts as a unitary rotation of u and d states,

$$\begin{pmatrix} u \\ d \end{pmatrix} \rightarrow U \begin{pmatrix} u \\ d \end{pmatrix}, \tag{17.1}$$

should be a symmetry of the strong interactions. Indeed, both in nuclear and in elementary particle physics, the quantum states form multiplets of this $SU(2)$ symmetry, called *isotopic spin* or *isospin*. Similarly, since the strange quark is only a little heavier than the u and d quarks, it makes sense to consider the symmetry of unitary transformations of the triplet (u, d, s). Gell-Mann and Ne'eman showed that the elementary particles naturally fill out irreducible representations of this $SU(3)$ symmetry.

Despite the phenomenological success of the original quark model, it had two serious problems. First, despite considerable effort, free particles with fractional charge could not be found. Second, the spectrum of baryons required the assumption that the wavefunction of the three quarks be totally symmetric under the interchange of the quark spin and flavor quantum numbers, contradicting the expectation that quarks, which must have spin 1/2, should obey Fermi-Dirac statistics. The need for this symmetry is most clearly illustrated in the fact that one of the lightest excited states of the nucleon is a spin-3/2 particle with charge $+2$, the Δ^{++}. This particle is readily interpreted as a uuu bound state with zero orbital angular momentum and all three quark spins parallel.

To reconcile the baryon spectrum with the spin-statistics theorem, Han and Nambu, Greenberg, and Gell-Mann proposed that quarks carry an additional, unobserved quantum number, called *color*. They introduced the *ad hoc* assumption that baryon wavefunctions must be totally antisymmetric in color quantum numbers. Then, if the quark wavefunctions are totally symmetric

in spin and flavor, they are totally antisymmetric overall, in agreement with Fermi-Dirac statistics. The simplest model of color would be to assign quarks to the fundamental representation of a new, internal $SU(3)$ global symmetry. Suppressing for a moment the spin and flavor quantum numbers, we can represent quarks by q_i, where $i = 1, 2, 3$ is the color index. Thus quarks transform under the fundamental, or "3", representation of the color $SU(3)$ symmetry. Antiquarks, $\bar{q}^i$, transform in the $\bar{3}$ representation. The inner product of a 3 and a $\bar{3}$ is an invariant of $SU(3)$. One can also form an invariant by using the totally antisymmetric combination of three 3's, ϵ_{ijk}: This object transforms under a unitary transformation according to

$$\epsilon_{ijk} \rightarrow U_{ii'}U_{jj'}U_{kk'}\epsilon_{i'j'k'} = (\det U)\epsilon_{ijk}, \tag{17.2}$$

and so is invariant under $SU(3)$ transformations, which have $\det U = 1$. Under the postulate that all hadron wavefunctions must be invariant under $SU(3)$ symmetry transformations, these two types of combinations are the only simple ones allowed:

$$\bar{q}^i q_i, \qquad \epsilon^{ijk} q_i q_j q_k, \qquad \epsilon_{ijk}\bar{q}^i\bar{q}^j\bar{q}^k. \tag{17.3}$$

That is, the assumption that physical hadrons are singlets under color implies that the only possible light hadrons are the mesons, baryons, and antibaryons.

Like the original quark model, the color hypothesis was phenomenologically successful but raised additional questions: Why should quarks have this seemingly superfluous property, and what mechanism insures that all hadron wavefunctions are color singlets? The answers to these questions came not from hadron spectroscopy, but from the deep-inelastic scattering experiments described in Chapter 14 and the ensuing search for a theory of parton binding with the property of asymptotic freedom. When it was discovered that non-Abelian gauge theories have this property, all that remained was to identify the correct gauge group and fermion representation. Since the color symmetry had no other obvious physical role, it was natural to identify this symmetry with the gauge group, with the colors being the gauge quantum numbers of the quarks. This reasoning resulted in a model of the strong interactions as a system of quarks, of the various flavors, each assigned to the fundamental representation of the local gauge group $SU(3)$. The quanta of the $SU(3)$ gauge field are called *gluons*, and the theory is known as Quantum Chromodynamics, or QCD.

In this book, we will mainly investigate the properties of QCD in the high-energy regime, where the coupling constant has become small. However, we should point out that one can also study QCD in the regime of strong coupling, using an approximation scheme introduced by Wilson in which the continuum gauge theory is replaced by a discrete statistical mechanical system on a four-dimensional Euclidean lattice. Using this approximation, Wilson showed that, for sufficiently strong coupling, QCD exhibits *confinement of color*: The only finite-energy asymptotic states of the theory are those that are singlets of color $SU(3)$. Thus the *ad hoc* assumption that explains the

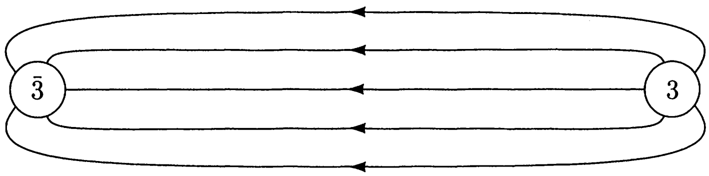


Figure 17.1. Gauge electric field configuration associated with the separation of color sources in a strong-coupling gauge theory.

spectrum of hadrons turns out to be a consequence of the non-Abelian gauge theory coupling to color. If one attempts to separate a color-singlet state into colored components—for example, to dissociate a meson into a quark and an antiquark—a tube of gauge field forms between the two sources, as shown in Fig. 17.1. In a non-Abelian gauge theory with sufficiently strong coupling, this tube has fixed radius and energy density, so the energy cost of separating color sources grows proportionally to the separation. A force law of this type can consistently be weak at short distances and strong at long distances, accounting for the fact that isolated quarks are not observed. We will discuss the large-distance, strong-coupling limit of gauge theories further in the Epilogue.

The short-distance limit of Quantum Chromodynamics can be readily studied using the Feynman diagram technology that we have developed in previous chapters. Here asymptotic freedom makes the coupling weak, and there is a sensible diagrammatic perturbation theory that begins from the model of free quarks and gluons. The following sections treat the elementary interactions among quarks and gluons that can be observed in high-energy experiments.

17.2 e^+e^- Annihilation into Hadrons

The simplest reaction involving quarks is the production of quark pairs in e^+e^- annihilation, a process that we treated already in Section 5.1. There we analyzed this process only at the most elementary level, viewing it as a pure QED reaction in which free quarks are created by a virtual photon. The diagram for this lowest-order process is shown in Fig. 17.2(a). The computation of the total cross section includes a sum over the various color states of the quark fields, and so provides a confirmation that the number of allowed colors is 3. Combining the color factor with the square of the quark electric charges, we found (Eq. (5.16))

$$\sigma(e^+e^- \to \text{hadrons}) = \sigma_0 \cdot 3 \cdot \sum_f Q_f^2, \tag{17.4}$$

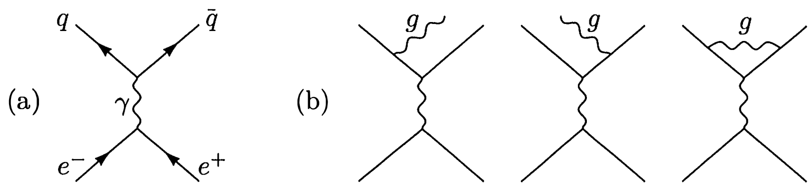


Figure 17.2. Diagrams contributing to the process $e^+e^- \to$ hadrons in QCD: (a) the leading-order diagram; (b) corrections of order α_s.

where σ_0 is the QED cross section for $e^+e^- \to \mu^+\mu^-$,

$$\sigma_0 = \frac{4\pi\alpha^2}{3s}, \tag{17.5}$$

and the sum in (17.4) is taken over quark flavors. This formula assumes that the center of mass energy is high enough that we can ignore the quark masses.

When we couple the quarks to an $SU(3)$ gauge theory, we add many important processes that affect both the value of this cross section and the final states that it includes. Some of the most important effects cannot be discussed within the context of perturbation theory. In particular, though the leading diagram contains free quarks, the particles that emerge from the reaction are color-singlet mesons and baryons. However, we will find that QCD perturbation theory with quarks and gluons does make a number of important predictions for e^+e^- annihilation to hadrons. The ideas that we develop in working out these predictions will also apply to many other strong-interaction processes.

Total Cross Section

The leading corrections to the rate of e^+e^- annihilation due to gluon exchange and emission are shown in Fig. 17.2(b). These are precisely the diagrams computed in the Final Project of Part I. The first two diagrams give a cross section of order g^2, where g is the $SU(3)$ gauge coupling, to produce a gluon in addition to the quark and antiquark. The third diagram must be summed in the amplitude with the leading diagram to produce a correction to the rate of $q\bar{q}$ production without gluon emission. In Part I, we computed these two contributions as if the strong interactions were an Abelian gauge theory. To obtain the corresponding expressions in QCD, we need only multiply the Abelian formulae by the group theory factor

$$\operatorname{tr}[t^a t^a] = C_2(r) \cdot \operatorname{tr}[1] = \frac{4}{3} \cdot 3, \tag{17.6}$$

where we have used Eq. (15.97) to evaluate $C_2(r)$ for the fundamental representation of $SU(3)$. The factor of 3 is the same color sum that appears in Eq. (17.4). Thus we can obtain the correct formulae for QCD from those of

the Final Project by making the replacement

$$g^2 \longrightarrow \frac{4}{3}g^2, \qquad \text{or} \qquad \alpha_g \longrightarrow \frac{4}{3}\alpha_s, \tag{17.7}$$

where

$$\alpha_s = \frac{g^2}{4\pi} \tag{17.8}$$

is the strong-interaction analogue of the fine-structure constant.

The end result of the Final Project of Part I was a formula for the total cross section to produce hadrons in e^+e^- annihilation. If we replace α_g with $(4/3)\alpha_s$, that result becomes

$$\sigma(e^+e^- \rightarrow \text{hadrons}) = \sigma_0 \cdot \Big(3\sum_f Q_f^2\Big) \cdot \Big[1 + \frac{\alpha_s}{\pi} + \mathcal{O}(\alpha_s^2)\Big]. \tag{17.9}$$

This is actually the sum of the rates for two elementary processes, $e^+e^- \rightarrow q\bar{q}$ (including the correction from the third diagram of Fig. 17.2(b)) and $e^+e^- \rightarrow q\bar{q}g$. Although the rate for each of these processes is divergent as the gluon mass is taken to zero, that divergence cancels when they are combined. This is another example of the phenomenon of infrared divergence cancellations that we studied for the example of electron scattering in Sections 6.4 and 6.5. There we showed that the dressing of the final state by the emission of soft and collinear photons does not affect the overall scattering rate. Here, we see again that infrared divergences cancel in the total rate, although the sum over real and virtual gluon corrections leaves over a simple numerical correction.

It is not difficult to understand the cancellation of infrared logarithms intuitively. The original process $e^+e^- \rightarrow q\bar{q}$ is extremely rapid: Since the virtual photon is off-shell by an amount $q^2 = s$, the quarks are created in a time $1/\sqrt{s}$. However, the emission of collinear gluons, and the virtual corrections associated with the exchange of soft gluons, occur over a much longer time scale. In the diagrams with gluon emission, the virtual quark or antiquark is off-mass-shell by an amount $p_{\perp g}^2$, where $p_{\perp g}$ is the transverse momentum of the gluon relative to the $q\bar{q}$ system. Thus this virtual state survives for a time $1/p_{\perp g}$ before it decays. Such a slow process cannot affect the probability that a $q\bar{q}$ pair was produced; it can only affect the properties of the final state into which the $q\bar{q}$ system will evolve. By this logic, the only perturbative corrections that can affect the total cross section are those for which $p_{\perp g} \sim \sqrt{s}$. Another way to express this conclusion would be to argue that, after contributions from the infrared-sensitive regions have canceled, the contributions that remain come from the region of large real or virtual gluon momenta. By either argument, formula (17.9) should be a meaningful prediction of QCD perturbation theory, even though it involves an integral over the region of soft gluon emission.

The Running of α_s

Formula (17.9) depends on the QCD coupling constant α_s, which must be defined at some renormalization point M. This is in contrast to the QED coupling constant, which we defined in a natural way by on-shell renormalization. In QCD, we would like to avoid discussing on-shell quarks, since these are strongly interacting particles that are significantly affected by nonperturbative forces. The use of an arbitrary renormalization point M allows us to avoid this problem. We will define α_s by renormalization conditions imposed at a large momentum scale M where the coupling is small; this value of α_s can then be used to predict the results of scattering processes with any large momentum transfer.

However, the use of renormalization at a scale M in a computation involving momentum invariants of order P^2 involves some subtlety when P^2 and M^2 are very different. In our discussion of Section 12.3, we saw that, in this circumstance, Feynman diagrams with n loops typically contain correction terms proportional to $(\alpha_s \log(P^2/M^2))^n$. Fortunately, we can absorb these corrections into the lowest-order terms by using the renormalization group to replace the fixed renormalized coupling with a running coupling constant.

To illustrate how this analysis applies to QCD, let us examine the implications of the Callan-Symanzik equation for the e^+e^- annihilation total cross section σ, viewed as a function of s, a renormalization scale M, and the value of α_s at the renormalization scale. Like the QED potential (12.87), the e^+e^- total cross section is an observable quantity and so its normalization is independent of any conventions. It therefore obeys a Callan-Symanzik equation with $\gamma = 0$:

$$\left[M\frac{\partial}{\partial M} + \beta(g)\frac{\partial}{\partial g}\right]\sigma(s, M, \alpha_s) = 0. \tag{17.10}$$

By dimensional analysis, we can write

$$\sigma = \frac{c}{s} f\left(\frac{s}{M^2}, \alpha_s\right), \tag{17.11}$$

were c is a dimensionless constant. Then the Callan-Symanzik equation implies that f depends on its arguments only through the running coupling constant $\alpha_s(Q) = \bar{g}^2/4\pi$, evaluated at $Q^2 = s$. The coupling constant $\bar{g}$ is defined to satisfy the renormalization group equation

$$\frac{d}{d\log(Q/M)}\bar{g} = \beta(\bar{g}), \tag{17.12}$$

with initial condition $\alpha_s(M) = \alpha_s$. For QCD with three colors and n_f approximately massless quarks, the β function is given by Eq. (16.135):

$$\beta(g) = -\frac{b_0 g^3}{(4\pi)^2}, \qquad \text{with } b_0 = 11 - \frac{2}{3}n_f. \tag{17.13}$$

Then the solution of the renormalization group equation is

$$\alpha_s(Q) = \frac{\alpha_s}{1 + (b_0\alpha_s/2\pi)\log(Q/M)}. \tag{17.14}$$

The explicit dependence of σ on α_s can be found by matching the successive terms in the expansion of $f(\alpha_s(\sqrt{s}))$ to the terms in the perturbative expansion. To the order of the first corrections, we find simply

$$\sigma = \sigma_0 \cdot \Big(3\sum_f Q_f^2\Big) \cdot \Big[1 + \frac{\alpha_s(\sqrt{s})}{\pi} + +\mathcal{O}(\alpha_s^2(\sqrt{s}))\Big]. \tag{17.15}$$

Thus the Callan-Symanzik equation instructs us to replace the fixed renormalized coupling α_s with the running coupling constant $\alpha_s(Q)$, evaluated at $Q^2 = s$.

Because the fixed coupling α_s depends on the arbitrary renormalization point M, it is sometimes useful to remove it from our formulae completely. To do this, we define a mass scale conventionally called Λ (not to be confused with an ultraviolet cutoff!) satisfying

$$1 = g^2(b_0/8\pi^2)\log(M/\Lambda). \tag{17.16}$$

Then Eq. (17.14) can be rearranged into the form

$$\alpha_s(Q) = \frac{2\pi}{b_0 \log(Q/\Lambda)}. \tag{17.17}$$

This formula is the clearest expression of the statement that $\alpha_s(Q)$ becomes small as $(\log(Q))^{-1}$ for large Q. The momentum scale Λ is the scale at which α_s becomes strong as Q^2 is decreased.

Experimental measurements of the rate of this reaction and others yield a value of $\Lambda \approx 200$ MeV. QCD perturbation theory is valid only when Q is somewhat larger than this, say above $Q = 1$ GeV, where $\alpha_s(Q) \approx 0.4$. The strong interactions become strong at distances larger than $\sim 1/\Lambda$, which is roughly the size of the light hadrons.

Although the example of the e^+e^- annihilation cross section is especially simple, since it depends on only one momentum invariant, similar conclusions carry over to other QCD predictions. In analyzing strong-interaction processes that are sensitive to the quark and gluon substructure, we will find leading-order formulae for the reaction cross sections that depend on the renormalized coupling α_s. To make these expressions satisfy the Callan-Symanzik equation, we must replace this fixed coupling with the running coupling constant $\alpha_s(Q)$, evaluated with Q of the order of the momentum invariants of the reaction. Since the running coupling constant depends only logarithmically on Q, we need not worry about choosing Q precisely. If we guess the proper scale incorrectly by a factor of 2, this induces an error in $\alpha_s(Q)$ that is of order $(\log(Q))^{-2} \sim \alpha_s^2(Q)$. Conversely, this ambiguity would be resolved by computing to the next order in α_s.

Before concluding this formal treatment of the e^+e^- annihilation cross section, we should add one qualification. At the beginning of Section 12.2, we remarked that renormalization group predictions can be complicated by the appearance of physical thresholds and their associated singularities, and so we stated these predictions only for when the relevant momentum invariant P^2 was large and spacelike. In the present chapter, we will be concerned with cross sections for quark and gluon reactions, evaluated on-shell. This introduces additional complications of principle. For example, in order to apply the Callan-Symanzik equation to $\sigma(s)$, we needed to know that this quantity contains no infrared divergences whose regulator might provide another dimensionful parameter. Throughout this chapter we will assume that similiar cancellations of divergences associated with soft and collinear gluons occur in the processes of interest to us. The complete proof of these cancellations in QCD can be carried through, but it is rather technical.* In some cases, an alternative point of view is possible; one can justify the use of the renormalization group to analyze an on-shell amplitude by relating it to Green's functions evaluated in the spacelike region. This method of analysis, which brings its own insights, will be the main subject of Chapter 18.

Gluon Emission and Jet Production

A second result of the Final Project of Part I was a formula for the differential cross section for $q\bar{q}$ production with gluon emission. Transcribing this formula to QCD using (17.7) gives the following result: Let x_1, x_2, x_3 be the ratios of the quark, antiquark, and gluon energies to the electron beam energy. These satisfy $0 < x_i < 1$ and $x_1+x_2+x_3 = 2$. Then the cross section for $e^+e^- \to q\bar{q}g$ is given by

$$\frac{d\sigma}{dx_1 dx_2}(e^+e^- \to q\bar{q}g) = \sigma_0 \cdot \Big(3\sum_f Q_f^2\Big) \cdot \frac{2\alpha_s}{3\pi}\frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}. \tag{17.18}$$

This cross section is singular as x_1 or x_2 approaches 1. The limit $x_1 \to 1$ corresponds to configurations in which the quark has the maximum possible energy, while the antiquark and the gluon go off in the opposite direction, sharing the remaining energy. Then the antiquark and gluon have almost collinear lightlike momentum vectors and so form a system of very small invariant mass. Similarly, the limit $x_2 \to 1$ corresponds to configurations in which the quark and gluon are collinear. These singularities are responsible for the divergence of the integrated cross section in the limit of vanishing gluon mass.

How should we interpret these singularities? In our general treatment of bremsstrahlung in Section 6.1, we saw that the emission of a photon by

*For a review of the theorems justifying the formulae of perturbative QCD, see J. C. Collins and D. E. Soper, *Ann. Rev. Nucl. Sci.* **37**, 383 (1987).

a scattered electron is enhanced, for collinear radiation, by a factor of order $\log(q^2/m^2)$, where m is the mass of the electron. Thus the total rate for emitting a collinear photon formally diverges in the limit of zero mass. The same conclusion holds for the emission of gluons by quarks. A divergence that appears for collinear emissions in the limit of zero mass is called a *mass singularity.* In QED, we saw that the mass singularity signals a real physical effect of strong collinear radiation when q^2 is large. In QCD, we might expect strong gluon radiation in this limit, but we must think carefully about how this radiation appears experimentally. Whether a collinear gluon is radiated or not, the quark and antiquark emerging from the reaction will undergo further soft interactions with the other products. These processes must continue, producing quark-antiquark pairs and emitting and absorbing gluons, until all colored particles are collected into color-singlet hadrons. Thus the presence of one or more collinear gluons will have no noticeable effect on the final state, which consists of two back-to-back jets of hadrons. For this reason, formula (17.18) is of no use when the gluon transverse momentum is less than the typical scale of soft gluon interactions, roughly 1 GeV.

When the gluon is emitted with substantial transverse momentum with respect to the $q\bar{q}$ axis, however, it is not possible for subsequent soft exchanges to recall or reverse this transverse momentum. In this case, the $q\bar{q}g$ system evolves into a system of three distinct jets of hadrons. Thus, sufficiently far from the collinear regions, we can interpret Eq. (17.18) as the cross section for producing events with three hadronic jets having energies x_1, x_2, x_3 times the electron beam energy.

By an analysis similar to that given above for the total cross section, we can improve Eq. (17.18) by replacing the fixed coupling constant α_s with a running $\alpha_s(Q)$. A reasonable choice for Q is the transverse momentum of the gluon, $p_{\perp g}$, described below Eq. (17.9). If this transverse momentum is too small, however, $\alpha_s(Q)$ will be large, and our leading-order formula will break down. This is a second reason why we cannot use formula (17.18) when the transverse momentum of the gluon is less than about 1 GeV.

On the other hand, when the gluon transverse momentum is much larger than 1 GeV, there is no reason to distrust QCD perturbation theory. Soft processes cannot disturb the three-jet nature of the hadronic state, and asymptotic freedom insures that the coupling constant is small, so that the leading order of perturbation theory will be a good approximation.

The three-jet cross section (17.18) is a good example of the type of prediction that one obtains from the use of perturbation theory in QCD. We describe a strong-interaction process by the invariant momentum transfer Q it gives to hadronic constituents. QCD perturbation theory makes predictions about the flow of energy and momentum in such a reaction into the final system of jets of hadrons. If Q is small, perturbation theory is invalid, and we obtain no useful prediction. However, if Q is large, the asymptotic freedom of QCD implies that Feynman diagrams for quarks and gluons will correctly predict the behavior of the final system of hadronic jets.

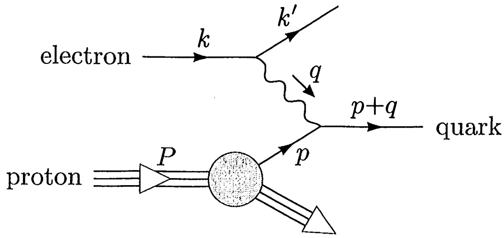


Figure 17.3. Deep inelastic scattering in QCD. The diagram shows the flow of momentum when a high-energy electron scatters from a quark taken from the wavefunction of the proton.

17.3 Deep Inelastic Scattering

After e^+e^- annihilation into hadrons, the next simplest reaction involving strongly interacting particles is electron scattering from a proton, or from some other hadron. At the most elementary level, this reaction can be viewed as the electromagnetic scattering of an electron from a quark inside the proton.† A way to visualize the process is shown in Fig. 17.3. Call the proton momentum P, and the initial quark momentum p. Call the initial and final momenta of the electron k and k'. If the final electron momentum is measured, one can deduce the momentum $q = k - k'$ transferred by the virtual photon to the hadronic system. The vector q is spacelike, and one conventionally writes $q^2 = -Q^2$.

If the invariant momentum transfer Q^2 is large, the quark is ejected from the proton in a manner that cannot be balanced by subsequent soft processes. These soft processes will, however, create gluons and quark-antiquark pairs that eventually neutralize the color and cause the struck quark to materialize as a jet of hadrons in the direction of the momentum transfer from the electron. Typically the total invariant mass of the final hadronic system is large, since the struck quark carries a large momentum with respect to the other "spectator" quarks. In this case, the process is referred to as *deep inelastic* scattering.

To derive a first approximation to the cross section for electron-proton scattering, we consider this reaction from a frame in which the electron and proton are moving rapidly toward each other, for example, the electron-proton center-of-mass frame. We assume that the center-of-mass energy is large enough that we can ignore the proton mass in working out the kinematics. Then the proton has an almost lightlike momentum along the collision axis. The constituents of the proton also have lightlike momenta, which

†The electron could just as well be a muon; all the same formulae apply in this case. Leptons can also scatter from quarks via the neutral-current weak interaction, as we will see in Chapter 20.

are almost collinear with the momentum of the proton. This is because a constituent cannot acquire a large transverse momentum except through exchange of a hard gluon, a process that is suppressed by the smallness of α_s at large momentum scales. Thus, to leading order in QCD perturbation theory, we can write

$$p = \xi P, \tag{17.19}$$

where ξ is a number between 0 and 1, called the *longitudinal fraction* of the constituent. To leading order in α_s we can also ignore gluon emission or exchanges during the collision process. The cross section for electron-proton scattering is then given by the cross section for electron-quark scattering at given longitudinal fraction ξ, multiplied by the probability that the proton contains a quark at that value of ξ, integrated over ξ.

But the probability that the proton contains a certain constituent with a certain momentum fraction cannot be computed using QCD perturbation theory, since it depends on the soft processes that determine the structure of the proton as a bound state of quarks and gluons. We will therefore consider this probability to be an unknown function, to be determined from experiment. Eventually, we will need to make use of such a probability function for each species of quark, antiquark, and gluon that can be found in the wavefunction of the proton. Collectively, these constituents are called *partons.* For each parton species f, we write the probability of finding a constituent of the proton of type f at longitudinal fraction ξ as

$$\begin{pmatrix} \text{probability of finding constituent } f \\ \text{with longitudinal fraction } \xi \end{pmatrix} = f_f(\xi) d\xi. \tag{17.20}$$

The functions $f_f(\xi)$ are called the *parton distribution functions.* Using this notation, the cross section for electron-proton inelastic scattering is given to leading order in α_s by the expression

$$\begin{aligned} \sigma\big(e^-(k)p(P) &\to e^-(k') + X\big) \\ &= \int_0^1 d\xi \sum_f f_f(\xi)\, \sigma\big(e^-(k) q_f(\xi P) \to e^-(k') + q_f(p')\big), \end{aligned} \tag{17.21}$$

where X stands for any hadronic final state. The sum in (17.21) contains contributions from constituent antiquarks as well as constituent quarks.

Equation (17.21) is equivalent to the formula (14.8) that we constructed for this reaction in Chapter 14. Now we see that this formula is justified by the smallness of the QCD coupling constant at large momentum scales. It is important to remember, however, that (17.21) is not the complete prediction of QCD, but only the first term of an expansion in α_s; this level of approximation is called the *parton model.* The higher-order QCD corrections to Eq. (17.21) will involve modifications both to the electron-quark scattering cross section and to the parton distribution functions. The most important of these corrections are discussed in Section 17.5.

In the same way, all other reactions of the proton that involve large momentum transfer also have parton model descriptions. In QCD, all of these reaction cross sections are computed from scattering amplitudes for quarks and gluons. The initial motion of the partons for any process is described by the same parton distribution functions $f_f(\xi)$ that appear in deep inelastic scattering.

Let us now work out the explicit leading-order formula for the deep inelastic scattering cross section, reviewing the analysis in Chapter 14. In Eq. (14.3), we wrote the leading-order differential cross section for the parton-level process,

$$\frac{d\sigma}{d\hat{t}}(e^- q \to e^- q) = \frac{2\pi\alpha^2 Q_f^2}{\hat{s}^2}\left[\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}\right]. \tag{17.22}$$

In general, we will use the symbols $\hat{s}$, $\hat{t}$, $\hat{u}$ to denote the Mandelstam variables for two-body scattering processes at the parton level. These variables must be related to observable properties of the hadronic system or the scattered electron. For massless initial and final particles,

$$\hat{s} + \hat{t} + \hat{u} = 0.$$

In the case of deep inelastic scattering,

$$\hat{t} = -Q^2$$

and

$$\hat{s} = 2p \cdot k = 2\xi P \cdot k = \xi s.$$

Thus the cross section for deep inelastic scattering at fixed Q^2 is given by

$$\frac{d\sigma}{dQ^2} = \int_0^1 d\xi \sum_f f_f(\xi) Q_f^2 \frac{2\pi\alpha^2}{Q^4}\left[1 + \left(1 - \frac{Q^2}{\xi s}\right)^2\right]\theta(\xi s - Q^2). \tag{17.23}$$

The final factor expresses the kinematic constraint $\hat{s} \geq |\hat{t}|$. Expression (17.23) should be an accurate first approximation to the deep inelastic scattering cross section when Q^2 is large. In that case, the corrections to this formula from hard gluon emissions and exchanges will be of order $\alpha_s(Q^2)$.

We also showed in Chapter 14 that the measurement of the scattered electron momentum k' and thus the momentum transfer q uniquely determines an allowed value of ξ for elastic electron-quark scattering. This value is given by Eq. (14.7):

$$\xi = x, \qquad \text{where} \quad x \equiv \frac{Q^2}{2P \cdot q}. \tag{17.24}$$

When (17.23) is expressed as a doubly differential cross section in x and Q^2, it becomes the simple product of a parton-level cross section and a sum of parton distribution functions evaluated at $\xi = x$. In the literature, the symbol x is often used interchangably with ξ, and we will follow this practice from here on.

It is especially convenient to represent the cross section in terms of dimensionless combinations of kinematic variables. One of these should be x; a good choice for the other is

$$y \equiv \frac{2P \cdot q}{2P \cdot k} = \frac{2P \cdot q}{s}. \tag{17.25}$$

In the frame in which the proton is at rest,

$$y = \frac{q^0}{k^0}, \tag{17.26}$$

that is, y is the fraction of the incident electron's energy that is transferred to the hadronic system. On the other hand, since $p = \xi P$, we can evaluate y in terms of parton variables:

$$y = \frac{2p \cdot (k - k')}{2p \cdot k} = \frac{\hat{s} + \hat{u}}{\hat{s}}, \tag{17.27}$$

so that

$$\frac{\hat{u}}{\hat{s}} = -(1 - y). \tag{17.28}$$

From Eq. (17.26) or (17.28), we see that $y \le 1$. The kinematically allowed region in the (x, y) plane is thus

$$0 \le x \le 1, \qquad 0 \le y \le 1.$$

To express Eq. (17.23) in terms of x and y, we need the formula

$$Q^2 = xys, \tag{17.29}$$

which follows from Eqs. (17.24) and (17.25), and the change of variables

$$d\xi\, dQ^2 = dx\, dQ^2 = \frac{dQ^2}{dy} dx\, dy = xs\, dx\, dy.$$

Then the differential cross section becomes

$$\frac{d^2\sigma}{dxdy}(e^- p \to e^- X) = \Big(\sum_f x f_f(x) Q_f^2\Big) \frac{2\pi\alpha^2 s}{Q^4} \left[1 + (1 - y)^2\right]. \tag{17.30}$$

The factor $1/Q^4$ comes from the square of the virtual photon propagator. Once this factor is removed, the dependence on x and y completely factorizes. Each half of this relation contains physical information. The fact that the parton distributions $f_f(x)$ depend only on x and are independent of Q^2 is the statement of Bjorken scaling. This tells us that the initial distribution of partons is independent of the details of the hard scattering. The y dependence of the cross section comes from the underlying parton cross section. In Chapter 5, we saw that the elementary QED cross sections, viewed in the high-energy limit, reflect the helicities of the interacting particles. The behavior $[1 + (1 - y)^2]$ in (17.30) is known as the *Callan-Gross relation*; it is specific

to the scattering of electrons from massless fermions. This relation gave evidence that the partons involved in deep inelastic scattering were fermions, at a time when the relation between partons and quarks was still unclear.

Deep Inelastic Neutrino Scattering

Because the sum over quark flavors factorizes in Eq. (17.30), one cannot determine the individual parton distribution functions $f_f(x)$ from electron scattering experiments alone. One can, however, obtain more detailed information on the structure of the proton through deep inelastic *neutrino* scattering.

Neutrinos have zero electric charge and so do not interact by photon exchange, but they do interact with quarks through weak interactions. We will discuss the weak interactions in detail in Chapter 20; for the moment, let us adopt a simplified description that concentrates on the elementary process shown in Fig. 17.4. In this process, a neutrino converts to the associated charged lepton,‡ exchanging a virtual massive vector boson, the W^+. This boson couples to a quark current that converts a d quark to a u quark. The effect of this exchange process is to provide a different, but completely characterized, method for injecting a hard momentum transfer q. The amplitude for this process is described by the effective Lagrangian

$$\Delta\mathcal{L} = \frac{g^2}{2}\frac{1}{m_W^2}\left[\bar{\ell}\gamma^\mu\left(\frac{1-\gamma^5}{2}\right)\nu\right]\left[\bar{u}\gamma_\mu\left(\frac{1-\gamma^5}{2}\right)d\right] + \text{h.c.}, \tag{17.31}$$

where ℓ, ν, d, u are the fermion fields associated with the charged lepton, the neutrino, and the d and u quarks, and g is the weak interaction coupling constant. The factor $1/m_W^2$ comes from the W boson propagator, considered in the limit $q^2 \ll m_W^2$. The first two factors are often written in terms of the *Fermi constant* G_F, defined as

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}. \tag{17.32}$$

This constant gives the strength of the weak interactions at energies much less than m_W. The crucial property of the weak interactions, shown explicitly in (17.31), is that the W boson couples only to the left-handed helicity states of relativistic fermions. The deeper significance of this property will be discussed in Chapter 20.

For technical reasons, it is easiest to do neutrino deep inelastic scattering using muon neutrinos, which convert to muons after emitting a W boson. It is equally feasible to scatter muon antineutrinos from nuclear targets, and, as we will see, it is interesting to compare the effects of neutrinos and antineutrinos. Since the proton contains a small admixture of the heavier quarks s, c, these also give small contributions to neutrino deep inelastic scattering. However, we will ignore these contributions in our discussion.

‡There is also a *neutral-current* weak interaction in which the neutrino remains a neutrino; see Problem 20.4.

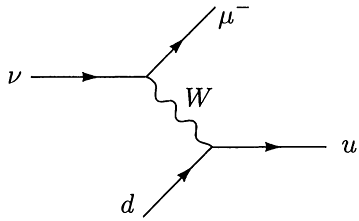


Figure 17.4. The elementary neutrino scattering process mediated by the weak interaction.

The cross section for neutrino deep inelastic scattering is given by a formula analogous to (17.21), with the photon-exchange cross section replaced by one resulting from W exchange. It is straightforward to work out this cross section directly. However, we can also obtain the result from Eq. (17.22), if we look back to Chapter 5 and recall how the structure of this equation arises from the various helicity contributions. In (17.22), the factor $\hat{t}^2$ in the denominator came from the photon propagator; this factor is replaced by m_W^4 in the weak interaction case. The factor $[\hat{s}^2 + \hat{u}^2]$ came from the Dirac matrix algebra. We saw in Section 5.2 that the first term is the contribution of left-handed electrons scattering from left-handed fermions or right-handed electrons scattering from right-handed fermions, and that the second term arises from the other helicity combinations. For the case of neutrino-quark scattering, the interaction (17.31) allows only the scattering of left-handed neutrinos from left-handed quarks, so only the $\hat{s}^2$ term appears. To determine the overall normalization of the cross section, we note that, since the neutrinos are produced in weak interactions, they always have left-handed polarization, so no polarization average should be done. On the other hand, we must still average over the polarization of the initial quark. In all, we find

$$\frac{d\sigma}{d\hat{t}}(\nu d \to \mu^- u) = \frac{\pi g^4}{2(4\pi)^2 \hat{s}^2}\left[\frac{\hat{s}^2}{m_W^4}\right] = \frac{G_F^2}{\pi}. \tag{17.33}$$

It is easy to check this formula by explicit computation starting from (17.31).

The cross section for the scattering of antineutrinos from quarks can be worked out in the same way. Note that this reaction involves the exchange of a W^-, and so converts u quarks to d quarks. However, the u quarks must still be left-handed. The only modification from the previous paragraph comes in the fact that the antineutrinos that couple to the interaction (17.31) are right-handed, so the cross section comes from the term in (17.22) proportional to $\hat{u}^2$:

$$\frac{d\sigma}{d\hat{t}}(\bar{\nu} u \to \mu^+ d) = \frac{\pi g^4}{2(4\pi)^2 \hat{s}^2}\left[\frac{\hat{u}^2}{m_W^4}\right] = \frac{G_F^2}{\pi}(1-y)^2. \tag{17.34}$$

Again, it is easy to verify this formula directly. The cross section for neutrino-antiquark scattering, converting a $\bar{u}$ into a $\bar{d}$, is also given by Eq. (17.34),

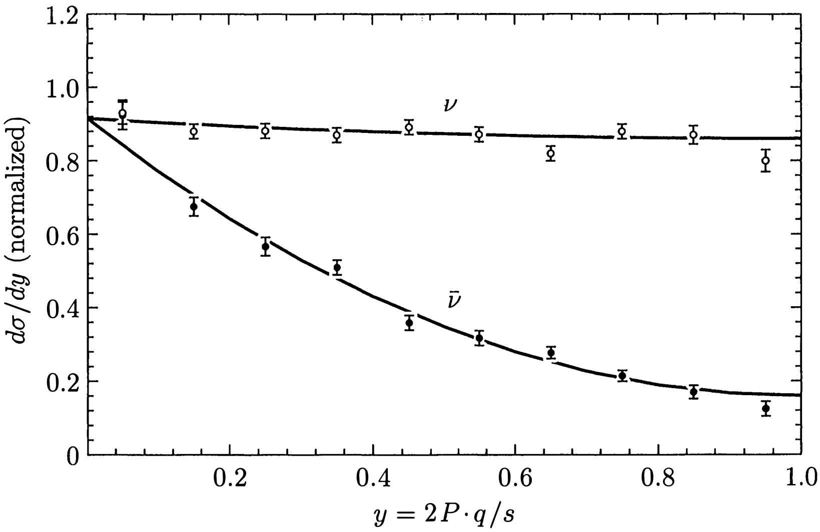


Figure 17.5. The distribution in y of neutrino and anti-neutrino deep inelastic scattering from an iron target, as measured by the CDHS experiment, J. G. H. de Groot, et. al., *Z. Phys.* **C1**, 143 (1979). The solid curves are fits to the form $A + B(1-y)^2$.

while the cross section for antineutrino-antiquark scattering, converting a $\bar{d}$ into a $\bar{u}$, is given by Eq. (17.33).

To convert these parton-level cross sections to physical cross sections, we combine them with the parton distribution functions. The kinematics is exactly the same as in the case of electron scattering. Thus, following the arguments that led to Eq. (17.30), we obtain the expressions

$$\begin{aligned} \frac{d^2\sigma}{dxdy}(\nu p \to \mu^- X) &= \frac{G_F^2 s}{\pi}\left[x f_d(x) + x f_{\bar{u}}(x)\cdot(1-y)^2\right], \\ \frac{d^2\sigma}{dxdy}(\bar{\nu} p \to \mu^+ X) &= \frac{G_F^2 s}{\pi}\left[x f_u(x)\cdot(1-y)^2 + x f_{\bar{d}}(x)\right]. \end{aligned} \tag{17.35}$$

According to these relations, deep inelastic neutrino scattering allows one to map separately the parton distribution functions for u and d quarks and antiquarks in the proton.

In addition, Eq. (17.35) makes a dramatic qualitative prediction: To the extent that a proton (or neutron) is made of quarks with very few additional quark-antiquark pairs, the deep inelastic neutrino scattering cross section should be constant in y, while the antineutrino scattering cross section should fall off as $(1-y)^2$. The measured y dependence of these deep inelastic cross sections is shown in Fig. 17.5. The qualitative behavior predicted by the parton description is clearly evident; the discrepancy from the strict prediction can be accounted for by a small antiquark component in the nucleon wavefunction.

The Parton Distribution Functions

Given that the parton model predictions for electron and neutrino deep inelastic scattering do fit the data, one can make use of these relations to extract the parton distribution functions and so learn something about the structure of the proton.* A set of distribution functions, chosen to fit all available data, is shown in Fig. 17.6. Since all of these distributions, especially those for antiquarks, peak sharply at small x, we have plotted $xf_f(x)$ for each species. As we remarked in Chapter 14, a small violation of Bjorken scaling is observed experimentally, so that these distribution functions change slowly with Q^2. The figure shows these functions at $Q^2 = 4\ \text{GeV}^2$. We will see in Section 17.5 that this violation of Bjorken scaling is an effect of higher-order corrections in QCD; we will also argue that the measurement of this scaling violation allows one to determine the parton distribution function for gluons, $f_g(x)$. Anticipating this result, we have also plotted this function in the figure. Not surprisingly, one finds that the u and d quarks are most likely to carry a substantial fraction of the proton's momentum, while antiquarks and gluons tend to have small longitudinal fractions.

Since the parton distributions are the probabilities of finding various proton constituents, they must be normalized in a way that reflects the quantum numbers of the proton. The proton is a bound state of uud, plus some admixture of quark-antiquark pairs. Thus it should contain an excess of two u quarks and one d quark over the corresponding antiquarks. These considerations imply the constraints

$$\int_0^1 dx\left[f_u(x) - f_{\bar{u}}(x)\right] = 2, \qquad \int_0^1 dx\left[f_d(x) - f_{\bar{d}}(x)\right] = 1. \tag{17.36}$$

So far we have discussed the parton distributions only for the proton. Similar considerations, however, apply to any other hadron. Each hadron has its own set of parton distribution functions; these obey sum rules analogous to (17.36) but reflecting the particular quantum numbers of the hadron. The parton distribution functions should also reflect the symmetries that link different hadrons. For example, since the neutron can be generated, to a few percent accuracy, by interchanging the role of u and d quarks in the proton, its distribution functions obey

$$f_u^n(x) = f_d(x), \qquad f_d^n(x) = f_u(x), \qquad f_{\bar{u}}^n(x) = f_{\bar{d}}, \qquad \text{etc.} \tag{17.37}$$

In these equations, and henceforth, a distribution function without a special label refers to the proton. The parton distribution functions of the antiproton are given by the exact relations

$$f_u^{\bar{p}}(x) = f_{\bar{u}}(x), \qquad f_{\bar{u}}^{\bar{p}}(x) = f_u(x), \qquad etc. \tag{17.38}$$

*A detailed discussion of the extraction of parton distribution functions from data can be found in G. Sterman, et. al., *Rev. Mod. Phys.* **67**, 157 (1995).

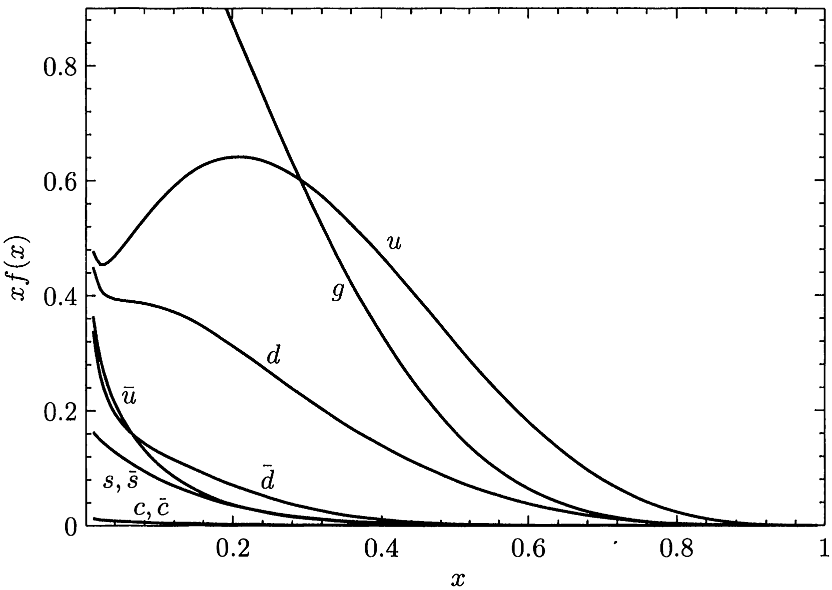


Figure 17.6. Parton distribution functions $xf_f(x)$ for quarks, antiquarks, and gluons in the proton, at $Q^2 = 4\ \mathrm{GeV}^2$. These distributions are obtained from a fit to deep inelastic scattering data performed by the CTEQ collaboration (CTEQ2L), described in J. Botts, et. al., *Phys. Lett.* **B304**, 159 (1993).

In any case, the total amount of momentum carried by the partons must sum to the total momentum of the hadron. This implies

$$\int_0^1 dx\, x\big[f_u(x) + f_d(x) + f_{\bar{u}}(x) + f_{\bar{d}}(x) + f_g(x)\big] = 1. \tag{17.39}$$

The distribution functions of quarks and antiquarks in the proton, as extracted from deep inelastic scattering data, contribute only about *half* of the total value required for this integral. The remaining energy-momentum must be carried by the gluons.

17.4 Hard-Scattering Processes in Hadron Collisions

If one collides hadrons with other hadrons at very high energy, most of the collisions will involve only soft interactions of the consituent quarks and gluons. Such interactions cannot be treated using perturbative QCD, because α_s is large when the momentum transfer is small. In some collisions, however, two quarks or gluons will exchange a large momentum $p_\perp$ perpendicular to the collision axis. Then, as in deep inelastic scattering, the elementary interaction takes place very rapidly compared to the internal time scale of the hadron wavefunctions, so the lowest-order QCD prediction should accurately

describe the process. Again, we should find a parton-model formula that is built from a leading-order subprocess cross section, integrated with parton distribution functions. For the case of proton-proton scattering, these functions will be the same ones that are measured in lepton-proton deep inelastic scattering.

For example, if the hard parton-level process involves quark-antiquark scattering into a final state Y, the leading-order QCD prediction takes the form

$$\begin{aligned}\sigma(p(P_1) + p(P_2) \to Y + X)& \\ = \int_0^1 dx_1 \int_0^1 dx_2 \sum_f f_f(x_1) f_{\bar{f}}(x_2) \cdot \sigma(q_f(x_1 P) + \bar{q}_f(x_2 P) \to Y),&\end{aligned} \tag{17.40}$$

where the sum runs over all species of quarks and antiquarks—u, d, $\bar{u}$, $\bar{d}$, ... (Here again, X denotes any hadronic final state.) The same formula, with appropriately modified distribution functions, applies to any other hadron-hadron collision. This formula will be a good first approximation if, by some invariant measure, a large momentum is transfered in the $q\bar{q}$ reaction. In this section we will discuss several examples of processes of this type.

Lepton Pair Production

The simplest example to analyze is the reaction in which a high-mass lepton pair $\ell^+\ell^-$ emerges from $q\bar{q}$ annihilation in a proton-proton collision. This reaction, called the *Drell-Yan process*, is illustrated in Fig. 17.7. In this case, the underlying $q\bar{q}$ reaction is described by an elementary QED cross section. To the leading order in QCD, the cross section that we require, $\sigma(q\bar{q} \to \ell^+\ell^-)$, is simply related to the cross section $\sigma(e^+e^- \to q\bar{q})$ given in (17.4). The only difference between the two calculations is that we must average rather than sum over the color orientations of the quark and antiquark. This gives two extra factors of 1/3. Thus,

$$\sigma(q_f \bar{q}_f \to \ell^+\ell^-) = \frac{1}{3} Q_f^2 \cdot \frac{4\pi\alpha^2}{3\hat{s}}. \tag{17.41}$$

If both final-state lepton momenta are observed, it is possible to reconstruct the total 4-momentum q of the virtual photon. It is also possible to determine the longitudinal fractions of the initial quark and antiquark, as we will now show. Let

$$M^2 = q^2 \tag{17.42}$$

be the square of the invariant mass of the Drell-Yan pair. (Do not confuse this quantity M with the renormalization scale.) If the initial partons have small transverse momentum, the transverse momentum of the virtual photon will also be small. Its longitudinal momentum, however, will in general be

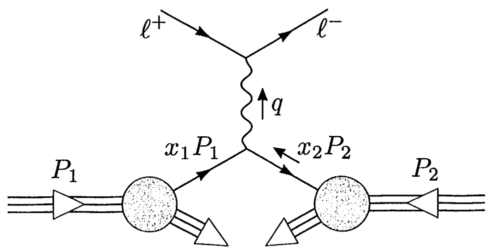


Figure 17.7. The Drell-Yan process: $pp \to \ell^+\ell^- +$ anything.

substantial. We parametrize this using the *rapidity*, Y, of the virtual photon, as defined in Eq. (3.48):

$$q^0 = M \cosh Y, \tag{17.43}$$

where q^0 is measured in the pp center of mass frame. We will express the longitudinal fractions of the quarks, and hence the Drell-Yan cross section, in terms of the observables M^2 and Y.

In the pp center of mass frame, the proton momenta take the explicit forms

$$P_1 = (E, 0, 0, E), \qquad P_2 = (E, 0, 0, -E),$$

where E satisfies $s = 4E^2$. Ignoring their small transverse momenta, we can write the constituent quark and antiquark momenta as x_1 and x_2 times these vectors, so that

$$q = x_1 P_1 + x_2 P_2 = \big((x_1+x_2)E, 0, 0, (x_1-x_2)E\big). \tag{17.44}$$

By computing the invariant square of this vector we find

$$M^2 = x_1 x_2 s. \tag{17.45}$$

Similarly, comparing (17.43) with (17.44), we find

$$\cosh Y = \frac{x_1 + x_2}{2\sqrt{x_1 x_2}} = \frac{1}{2}\left(\sqrt{\frac{x_1}{x_2}} + \sqrt{\frac{x_2}{x_1}}\right),$$

which implies

$$\exp Y = \sqrt{\frac{x_1}{x_2}}. \tag{17.46}$$

These equations can be inverted to determine x_1 and x_2:

$$x_1 = \frac{M}{\sqrt{s}} e^{Y}, \qquad x_2 = \frac{M}{\sqrt{s}} e^{-Y}. \tag{17.47}$$

Relations (17.45) and (17.46) let us convert the integral in Eq. (17.40) into an integral over the parameters M^2, Y of the produced leptons. The Jacobian of the change of variables is

$$\frac{\partial(M^2, Y)}{\partial(x_1, x_2)} = \begin{vmatrix} x_2 s & x_1 s \\ 1/2x_1 & -1/2x_2 \end{vmatrix} = s = \frac{M^2}{x_1 x_2}.$$

The cross section for lepton pair production is therefore

$$\frac{d^2\sigma}{dM^2dY}(pp \to e^+e^- + X) = \sum_f x_1 f_f(x_1)\, x_2 f_{\bar{f}}(x_2) \cdot \frac{1}{3}Q_f^2 \cdot \frac{4\pi\alpha^2}{3M^4}, \qquad (17.48)$$

where x_1 and x_2 are given by Eq. (17.47). It is remarkable that the cross section for the Drell-Yan process is determined point by point by information derivable from deep inelastic scattering. Unfortunately, the relation between the two processes implied by (17.48) receives a correction of order $\alpha_s(M)$ that turns out to be numerically large, and which must be included to check this prediction against experiment.

General Kinematics of Pair Production

In deriving Eq. (17.48), we used the total cross section (17.41) for the parton-level process, integrated over the angular distribution of the outgoing leptons. In principle, we could have retained the angular information and derived a triply differential distribution. This would be the most complete prediction possible for a two-body parton-level reaction. It will be useful to work out the kinematics of such reactions, taking a more general viewpoint. In the generic situation, a parton of type 1 from proton 1 scatters from a parton of type 2 from proton 2, yielding partons of types 3 and 4, with a squared momentum transfer $\hat{t}$. This generic process is shown in Fig. 17.8. In the Drell-Yan process, partons 3 and 4 are leptons. But these partons could also be quarks or gluons, which materialize as hadronic jets. We assume that all partons can be treated as massless. In parton variables, the cross section for this process is

$$\frac{d^3\sigma}{dx_1dx_2d\hat{t}}(pp \to 3+4+X) = f_1(x_1)f_2(x_2)\frac{d\sigma}{d\hat{t}}(1+2 \to 3+4). \qquad (17.49)$$

Let us now translate this formula to observable parameters of the final state.

In the leading order of QCD, the transverse momenta of partons 3 and 4 must be equal and opposite, but their longitudinal momenta are not constrained. We will take the three parameters of the final state to be the common magnitude of the parton transverse momenta $p_\perp$ and the *longitudinal rapidities* y_3, y_4 of the final-state partons, defined by the formulae

$$E_i = p_\perp \cosh y_i; \qquad p_{i\parallel} = p_\perp \sinh y_i. \qquad (17.50)$$

The longitudinal rapidity y_i gives the boost of the particle i from the frame where it has zero longitudinal momentum.† Recall from Section 3.3 that rapidities simply add under collinear boosts. The transverse momentum is invariant under longitudinal boosts. Thus, $(y_3, y_4, p_\perp)$ is a set of variables with convenient Lorentz transformation properties with respect to boosts along the

†In the literature on hadron collisions, y_i is usually called simply the rapidity, with the restriction to longitudinal boosts being understood.

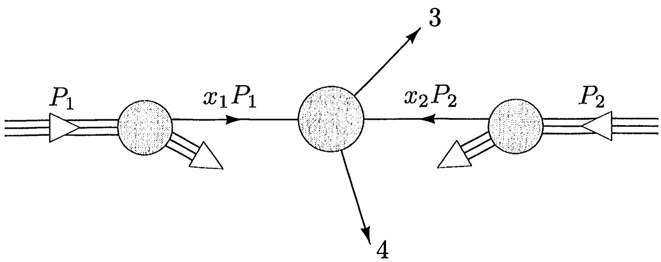


Figure 17.8. A generic two-body parton scattering process.

collision axis. We will now see that these three parameters are related in a straightforward way to the underlying variables x_1, x_2, $\hat{t}$.

Consider the center of mass frame of the colliding partons. The total energy in this frame is $\sqrt{\hat{s}}$. Let us use a subscript $*$ to denote other quantities measured in this frame, for instance, θ_* for the parton scattering angle. Then

$$p_{3\|*} = \tfrac{1}{2}\sqrt{\hat{s}}\cos\theta_*, \qquad p_{3\perp *} = \tfrac{1}{2}\sqrt{\hat{s}}\sin\theta_*, \tag{17.51}$$

and p_{4*} is oriented just oppositely. This frame is also the center of mass frame of partons 3 and 4, so

$$y_{3*} = -y_{4*} \equiv y_*. \tag{17.52}$$

Since rapidities transform by shifts, we can solve for y_* and for the rapidity Y by which one must boost to reach this frame:

$$y_* = \tfrac{1}{2}(y_3 - y_4), \qquad Y = \tfrac{1}{2}(y_3 + y_4). \tag{17.53}$$

The scattering angle θ_* is determined from y_* by combining (17.51) with the relation $E_* = p_\perp \cosh y_*$:

$$\frac{1}{\sin\theta_*} = \cosh y_*. \tag{17.54}$$

Then the Mandelstam variables

$$\hat{s} = \frac{4p_\perp^2}{\sin^2\theta_*}, \qquad \hat{t} = -\tfrac{1}{2}\hat{s}(1-\cos\theta_*) \tag{17.55}$$

can be expressed as

$$\hat{s} = 4p_\perp^2\cosh^2 y_*, \qquad \hat{t} = -2p_\perp^2\cosh y_*\, e^{-y_*}. \tag{17.56}$$

We can combine the first of these expressions with (17.47) to determine x_1 and x_2:

$$x_1 = \frac{2p_\perp}{\sqrt{s}}\cosh y_*\, e^{Y}, \qquad x_2 = \frac{2p_\perp}{\sqrt{s}}\cosh y_*\, e^{-Y}. \tag{17.57}$$

To translate the cross section (17.49) to the final parton observables, we need the Jacobian

$$\frac{\partial(x_1, x_2, \hat{t})}{\partial(y_3, y_4, p_\perp)} = \frac{8p_\perp^3}{s}\cosh^2 y_* = \frac{2p_\perp\hat{s}}{s}. \tag{17.58}$$

Multiplying Eq. (17.49) by this factor gives

$$\frac{d^3\sigma}{dy_3 dy_4 dp_{\perp}} = f_1(x_1) f_2(x_2)\,\frac{2p_\perp \hat{s}}{s}\,\frac{d\sigma}{d\hat{t}}(1+2\to 3+4). \qquad (17.59)$$

This can be simplified a bit using the relations $\hat{s} = x_1 x_2 s$ and $p_\perp dp_\perp = d^2 p_\perp / 2\pi$, yielding the final result:

$$\frac{d^4\sigma}{dy_3 dy_4 d^2 p_{\perp}} = x_1 f_1(x_1)\, x_2 f_2(x_2)\,\frac{1}{\pi}\frac{d\sigma}{d\hat{t}}(1+2\to 3+4). \qquad (17.60)$$

In this formula, x_1, x_2, and the Mandelstam variables of the parton subprocess are given by Eqs. (17.57) and (17.56).

This result gives us the complete distribution of final-state leptons or jets in any two-body reaction of partons. For example, to find the distribution of final-state leptons in the Drell-Yan process, we would insert into this formula the differential cross section for quark annihilation into leptons,

$$\frac{d\sigma}{d\hat{t}}(q_f \bar{q}_f \to \ell^+\ell^-) = \frac{1}{3}Q_f^2 \cdot \frac{2\pi\alpha^2}{\hat{s}^2}\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}. \qquad (17.61)$$

The formula can be applied equally well to other two-body parton reactions, if we know the relevant parton-level differential cross sections.

Jet Pair Production

The most common two-body parton reactions are those of QCD, involving quarks, gluons, or both. Unfortunately, it is very difficult to distinguish hadronic jets initiated by gluons from those initiated by quarks. It is even more difficult to determine experimentally whether the initial partons in a hard-scattering process were quarks or gluons. Thus, the predictions of QCD for hard-scattering processes are most often quoted as cross sections for jet production in hadronic collisions, summing over all possible reactions of quarks, antiquarks, and gluons. In any event, to derive these predictions, we must work out the basic parton-parton cross sections.

The simple two-body scattering processes of quarks, antiquarks, and gluons are the elementary processes of QCD perturbation theory, in the same sense that the reactions studied in Chapter 5 are the elementary processes of QED perturbation theory. They are the basic hadronic hard-scattering reactions that appear in QCD at the leading order in α_s. In the remainder of this section, we will write down the cross section formulae for the various possible quark and gluon subprocesses. All of these cross sections will be of order α_s^2. In practice, this α_s should be evaluated at a typical momentum transfer of the reaction, for example, at $Q^2 = \hat{t}$.

The simplest subprocess is the scattering of different species of quarks, for example, $u + d \to u + d$. At order α_s^2, this process occurs via the Feynman diagram shown in Fig. 17.9. This process is analogous to electron-muon

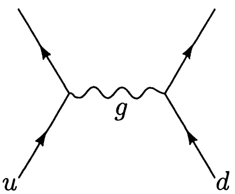


Figure 17.9. Feynman diagram contributing to $ud \to ud$.

scattering in QED, for which we wrote the cross section in Eq. (17.22):

$$\frac{d\sigma}{d\hat{t}}(e^-\mu \to e^-\mu) = \frac{2\pi\alpha^2}{\hat{s}^2}\left[\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}\right]. \tag{17.62}$$

To convert this to the cross section for quark scattering in QCD, we need only replace the QED coupling e^2 by g^2 times an $SU(3)$ group theory factor. The QCD diagram contains the factor

$$(t^a)_{i'i}(t^a)_{j'j},$$

where i, i' are the initial and final colors of the u quark and j, j' are the initial and final colors of the d quark. To compute the cross section, we must square this factor, sum over final colors, and average over initial colors. This gives the factor

$$\frac{1}{3}\cdot\frac{1}{3}\cdot \mathrm{tr}[t^b t^a]\cdot \mathrm{tr}[t^b t^a] = \frac{1}{9}\big[C(r)\big]^2\delta^{ab}\delta^{ab} = \frac{1}{9}\cdot\frac{1}{4}\cdot 8 = \frac{2}{9}, \tag{17.63}$$

where we have used Eq. (15.78) and $C(r) = 1/2$ for the fundamental representation of $SU(3)$. Thus for ud scattering,

$$\frac{d\sigma}{d\hat{t}}(ud \to ud) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}\right]. \tag{17.64}$$

The same formula applies for the scattering of any two different quarks, or, by crossing, to the scattering of a quark and an antiquark of a different species. Crossing from the t to the s channel gives the cross section for $q\bar{q}$ annihilation into a different species:

$$\frac{d\sigma}{d\hat{t}}(u\bar{u} \to d\bar{d}) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}\right]. \tag{17.65}$$

The scattering of a quark with an antiquark of the same species is more complicated, since now there are two Feynman diagrams, shown in Fig. 17.10, which interfere with one another. The analogous QED process is Bhabha scattering, $e^+e^- \to e^+e^-$, for which we worked out the cross section in Problem 5.2:

$$\frac{d\sigma}{d\hat{t}}(e^+e^- \to e^+e^-) = \frac{2\pi\alpha^2}{\hat{s}^2}\left[\left(\frac{\hat{s}}{\hat{t}}\right)^2 + \left(\frac{\hat{t}}{\hat{s}}\right)^2 + \hat{u}^2\left(\frac{1}{\hat{s}}+\frac{1}{\hat{t}}\right)^2\right]. \tag{17.66}$$

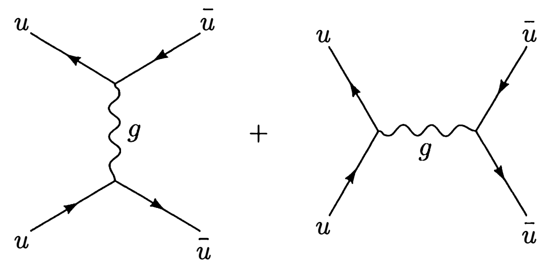


Figure 17.10. Feynman diagrams contributing to $u\bar{u} \to u\bar{u}$.

However, it is not quite straightforward to transcribe this to QCD, because different terms receive different color factors.

This process is most easily analyzed using initial and final states of definite helicity. For massless fermions, helicity is conserved, so the reaction $e^+_R e^-_L \to e^+_L e^-_R$ can receive a contribution only from the s-channel diagram, while $e^+_R e^-_R \to e^+_R e^-_R$ can receive a contribution only from the t-channel diagram. The corresponding cross sections are

$$\begin{aligned} \frac{d\sigma}{d\hat{t}}(e^+_R e^-_L \to e^+_L e^-_R) &= \frac{4\pi\alpha^2}{\hat{s}^2}\left(\frac{\hat{t}}{\hat{s}}\right)^2, \\ \frac{d\sigma}{d\hat{t}}(e^+_R e^-_R \to e^+_R e^-_R) &= \frac{4\pi\alpha^2}{\hat{s}^2}\left(\frac{\hat{s}}{\hat{t}}\right)^2. \end{aligned} \tag{17.67}$$

The cross section for $e^+_R e^-_L \to e^+_L e^-_L$ must vanish. The fourth possible process involving e^+_R receives contributions from both s- and t-channel diagrams. Computing this contribution explicitly, one finds

$$\frac{d\sigma}{d\hat{t}}(e^+_R e^-_L \to e^+_R e^-_L) = \frac{4\pi\alpha^2}{\hat{s}^2}\hat{u}^2\left(\frac{1}{\hat{t}} + \frac{1}{\hat{s}}\right)^2; \tag{17.68}$$

the cross term in the square is the interference term between the two diagrams. The invariance of QED under parity implies that the values of all of these cross sections remain identical when all helicities are reversed. It is easy to check that the spin-averaged cross section is indeed given by (17.66).

To convert Eq. (17.66) to a QCD cross section averaged over colors, we can assign the color factor (17.63) to the square of any individual diagram. However, the cross term between the two diagrams in Fig. 17.10 receives a different color factor:

$$\left(\frac{1}{3}\right)^2 \cdot (t^a)_{i'i}(t^a)_{jj'} \cdot (t^b)_{j'i'}(t^b)_{ij} = \frac{1}{9}\,\mathrm{tr}[t^a t^b t^a t^b]. \tag{17.69}$$

To evaluate this factor, we can make use of Eq. (16.79):

$$t^a t^b t^a t^b = \Big(C_2(r) - \frac{1}{2}C_2(G)\Big)t^a t^a = \left(\frac{4}{3} - \frac{3}{2}\right)\frac{4}{3} = -\frac{2}{9}.$$

So the color factor (17.69) equals $-2/27$.

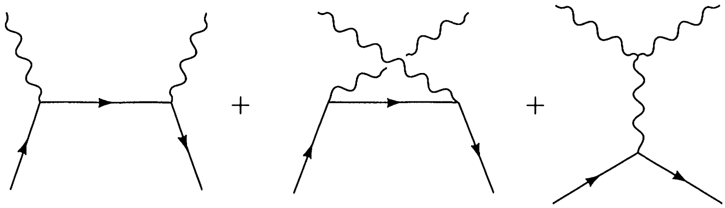


Figure 17.11. Feynman diagrams contributing to $q\bar{q} \to gg$.

Assembling the color factors and the helicity cross sections, we find the following result for the $u\bar{u}$ scattering cross section:

$$\frac{d\sigma}{d\hat{t}}(u\bar{u} \to u\bar{u}) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2} + \frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2} - \frac{2}{3}\frac{\hat{u}^2}{\hat{s}\hat{t}}\right]. \tag{17.70}$$

By crossing between the s and u channels, we find the corresponding cross section for $uu \to uu$:

$$\frac{d\sigma}{d\hat{t}}(uu \to uu) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[\frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2} + \frac{\hat{t}^2 + \hat{s}^2}{\hat{u}^2} - \frac{2}{3}\frac{\hat{s}^2}{\hat{u}\hat{t}}\right]. \tag{17.71}$$

The process $\bar{u}\bar{u} \to \bar{u}\bar{u}$ has the same cross section. This completes our catalogue of cross sections for the scattering of quarks and antiquarks.

We turn next to processes that involve both quarks and gluons. We will begin with the reaction $q\bar{q} \to gg$. This is the analogue of the QED annihilation of e^+e^- to $\gamma\gamma$, discussed in Section 5.5. The QED cross section is

$$\frac{d\sigma}{d\hat{t}}(e^+e^- \to \gamma\gamma) = \frac{2\pi\alpha^2}{\hat{s}^2}\left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}}\right]. \tag{17.72}$$

Since the photons are identical particles, this expression should be integrated over only half of the 4π solid angle.

The QCD reaction is considerably more complicated. As we saw in Section 16.1, this process receives contributions from three Feynman diagrams, shown in Fig. 17.11. These contributions must be summed over the transverse polarization states of the gluons. If one chooses instead to evaluate the sum over gluon polarizations by the replacement

$$\sum_{\epsilon} \epsilon^{\mu}\epsilon^{*\nu} \to -g^{\mu\nu}, \tag{17.73}$$

we saw in Section 16.3 that one must also include the (negative) cross section for $q\bar{q}$ annihilation to a ghost-antighost pair.

The leading behavior of the $q\bar{q} \to gg$ cross section as $\hat{t}$ or $\hat{u} \to 0$ is not so hard to evaluate. In either case, only the single diagram with the corresponding kinematic singularity contributes. The color factor associated with the square of either of these diagrams is the square of

$$(t^a)_{ij}(t^b)_{jk},$$

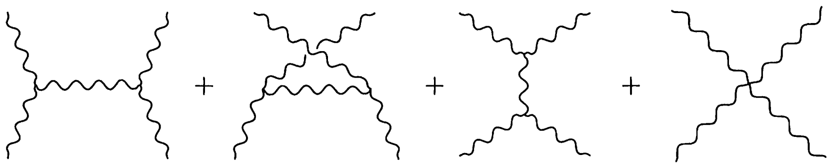


Figure 17.12. Feynman diagrams contributing to $gg \to gg$.

summed over the gluon colors a, b and averaged over the q and $\bar{q}$ colors i, k. This gives

$$\left(\frac{1}{3}\right)^2 \cdot \text{tr}[t^a t^b t^b t^a] = \frac{1}{9} \cdot 3\big(C_2(r)\big)^2 = \frac{16}{27}. \tag{17.74}$$

Thus the most singular terms are given by the QED result, with α replaced by α_s, multiplied by 16/27. The complete evaluation of the cross section is left for Problem 17.3; the result is

$$\frac{d\sigma}{d\hat{t}}(q\bar{q} \to gg) = \frac{32\pi\alpha_s^2}{27\hat{s}^2}\left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}} - \frac{9}{4}\left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\right)\right]. \tag{17.75}$$

The cross sections for the remaining quark-gluon processes can be obtained from this result by crossing. The result for the inverse reaction $gg \to q\bar{q}$ involves the same squared matrix element as (17.75); the only difference is that we average over gluon rather than quark colors, giving a relative factor of $(3/8)^2$. Thus,

$$\frac{d\sigma}{d\hat{t}}(gg \to q\bar{q}) = \frac{\pi\alpha_s^2}{6\hat{s}^2}\left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}} - \frac{9}{4}\left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2}\right)\right]. \tag{17.76}$$

For the reaction $qg \to qg$, cross the s and t channels in Eq. (17.75) and multiply by 3/8 since there is one gluon color average. This gives

$$\frac{d\sigma}{d\hat{t}}(qg \to qg) = \frac{4\pi\alpha_s^2}{9\hat{s}^2}\left[-\frac{\hat{u}}{\hat{s}} - \frac{\hat{s}}{\hat{u}} + \frac{9}{4}\left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2}\right)\right]. \tag{17.77}$$

The cross section for $\bar{q}g \to \bar{q}g$ is identical.

The final elementary process of QCD is gluon-gluon scattering. This has no QED analogue, and is rather tedious to evaluate. There are four leading-order diagrams, shown in Fig. 17.12. We discuss this process also in Problem 17.3. The final result for the spin- and color-averaged cross section is

$$\frac{d\sigma}{d\hat{t}}(gg \to gg) = \frac{9\pi\alpha_s^2}{2\hat{s}^2}\left[3 - \frac{\hat{t}\hat{u}}{\hat{s}^2} - \frac{\hat{s}\hat{u}}{\hat{t}^2} - \frac{\hat{s}\hat{t}}{\hat{u}^2}\right]. \tag{17.78}$$

The various parton cross sections listed in this section can be combined with the parton distribution functions to predict the cross section for jet production in hadron-hadron collisions. As an example, we show in Fig. 17.13 a comparison of the invariant mass ($\hat{s}$) distribution predicted for parton-parton

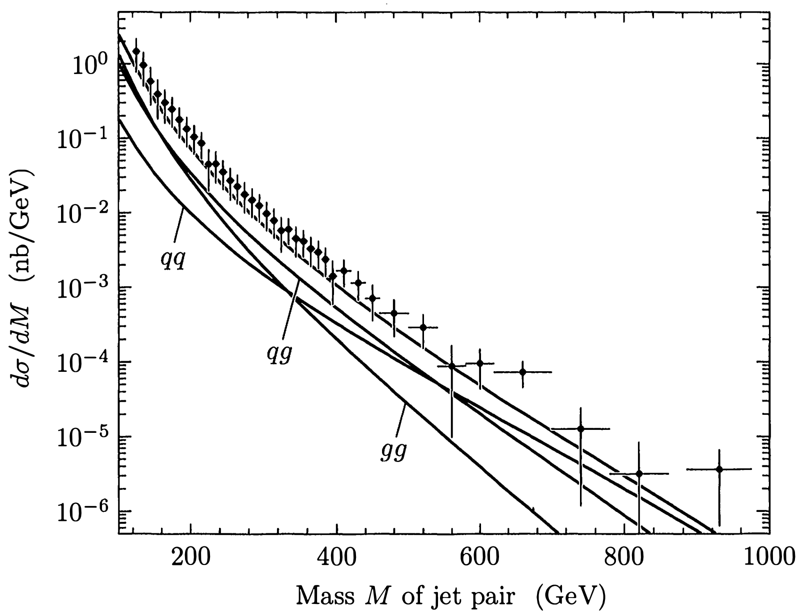


Figure 17.13. Two-jet invariant mass distribution in $p\bar{p}$ collisions at $E_{\rm cm}$ = 1.8 TeV, as measured by the CDF collaboration, F. Abe, et. al., *Phys. Rev.* **D48,** 998 (1993). The measurement is compared to a leading-order QCD calculation using the CTEQ structure functions described in Fig. 17.6. The three lower curves show the invariant mass distributions for the three components of the theoretical prediction: quark-quark (and antiquark) scattering, quark-gluon scattering, and gluon-gluon scattering.

scattering with the invariant mass distribution of two-jet events observed in high-energy $p\bar{p}$ collisions. The overall normalization of the theoretical prediction is uncertain by about a factor of 2 due to the ambiguity of the choice of Q^2 used to evaluate $\alpha_s(Q^2)$ in the parton cross sections, and due to similar ambiguities in deriving parton distributions from deep inelastic scattering cross sections. This uncertainty is reduced to about 30% when corrections of order α_s are included. Still, it is remarkable that the lowest-order QCD prediction tracks the observed distribution as a function of the two-jet invariant mass as it falls by six orders of magnitude. Thus, for the jet production cross section, as for hard processes involving leptons, QCD indeed gives a reasonable description of the behavior of the strong interactions at large momentum transfer.

17.5 Parton Evolution

Now that we have examined the predictions of QCD at the leading order for several strong interaction processes, we should investigate the corrections to these predictions at the next order in α_s. We saw in Section 17.2 that the corrections from individual diagrams may contain mass singularities, singularities associated with collinear emission processes which appear in the limit of zero mass. For the process of e^+e^- annihilation to hadrons, we saw that these mass singularities, and the infrared divergences from soft gluon emission, cancel in the expression for the total cross section. It can be shown that this is a general feature of processes in which quarks and gluons are produced in the collision of leptons or photons. However, when quarks or gluons appear in the initial state of a parton subprocess, the corrections to the process will, in general, have mass singularities that do not cancel. In this section we will demonstrate this effect and work out its physical interpretation. We will find that these singular terms predict a violation of Bjorken scaling by terms depending on the logarithm of the momentum scale. In fact, they lead to a precise set of differential equations that govern the momentum dependence of the parton distributions.

The basic phenomena associated with mass singularities in QCD are already present in the physics of collinear photon emission in QED at high energies, and so it is most straightforward to begin by studying that case. In this section, we will show that collinear photon emission leads to an analogue of a parton distribution function for the electron. We will derive a differential equation describing this distribution function, first constructed by Gribov and Lipatov. Finally, we will generalize this equation to QCD, following the construction of Altarelli and Parisi.[‡]

In Chapters 5 and 6, we studied several examples of QED processes that involved diagrams with t- or u-channel singularities. In these cases, we found that the total cross section was generally enhanced by an extra factor $\log(s/m^2)$ in the high-energy limit. For example, in Eq. (5.95) we saw that the u-channel exchange diagram in Compton scattering, Fig. 17.14(a), leads to an integral that, in the high-energy limit, takes the form

$$\int \frac{d\cos\theta}{(1+\cos\theta)}.$$

The singularity as $\cos\theta \to -1$ is cut off by the electron mass, leading to the logarithmic enhancement factor. Thus the collinear photon emission costs a factor that is not α but rather $\alpha\log(s/m^2)$. Emission of multiple collinear photons, as in Fig. 17.14(b), gives contributions of order $(\alpha\log(s/m^2))^n$. To improve the accuracy of perturbation theory, it would be useful to find a procedure for summing these terms to all orders in α. In QCD, the corresponding

[‡]V. N. Gribov and L. Lipatov, *Sov. J. Nucl. Phys.* **15**, 438 (1972); G. Altarelli and G. Parisi, *Nucl. Phys.* **B126**, 298 (1977). We also strongly recommend reading the papers of J. Kogut and L. Susskind, *Phys. Rev.* **D9**, 697, 3391 (1974).

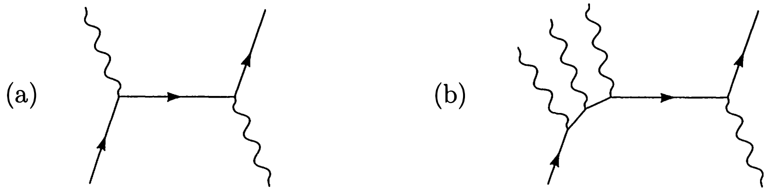


Figure 17.14. Diagrams with mass singularities associated with collinear photon emission: (a) leading order; (b) higher order.

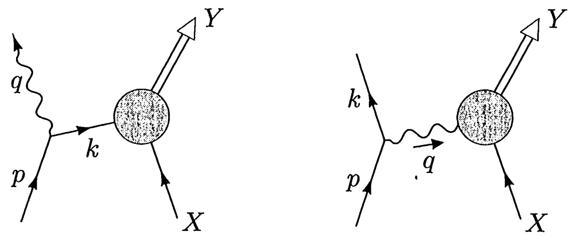


Figure 17.15. General form of diagrams with mass singularities in QED.

factor for collinear gluon emission would be

$$\alpha_s(Q^2)\log\frac{Q^2}{\mu^2},$$

where μ is the momentum scale where nonperturbative QCD effects become important. Comparing with Eq. (17.17), we see that this product is of order 1. Thus, in this case, the resummation of large logarithms is essential if we are to make any quantitative predictions.

In QED, diagrams with mass singularities associated with one collinear emission are of one of the forms shown in Fig. 17.15. In each case, the circle represents a scattering process with large momentum transfer. The mass singularity appears when the denominator of the intermediate propagator vanishes, that is, when the intermediate state is almost on-shell. Thus, it is natural to consider the first diagram in Fig. 17.15 to be a transition to a real photon and an almost-real electron, followed by the interaction of the electron with the remaining particles in the amplitude. The second diagram should have a similar interpretation with an almost-real photon in the intermediate state.

The only subtlety comes in defining the polarization of the intermediate-state particle. For the case of an intermediate-state electron, the numerator of the propagator is

$$\not{k} = \sum_s u^s(k)\bar{u}^s(k). \tag{17.79}$$

Thus, when $k^2 \to 0$, the photon emission vertex and the remaining part of the amplitude are contracted with on-shell polarization spinors for a massless

electron. The analogous statement for the diagram with the photon in the intermediate state would be that the electron emission vertex and the remaining photon amplitude should be contracted with physical transverse polarization vectors for the intermediate-state photon. Since the numerator of the photon propagator is $g^{\mu\nu}$, it is not obvious that the photon propagator reduces in this way. But it is true. To see this, use the expansion for $g^{\mu\nu}$ in terms of massless polarization vectors given in Eq. (16.20):

$$g^{\mu\nu} = \epsilon_+^\mu \epsilon_-^{\nu *} + \epsilon_-^\mu \epsilon_+^{\nu *} - \sum_i \epsilon_{Ti}^\mu \epsilon_{Ti}^{\nu *}. \tag{17.80}$$

Here ϵ_{Ti}^μ are transverse polarization vectors. The forward polarization vector ϵ_+^μ is proportional to the photon momentum q^μ. When we contract ϵ_+^μ with the QED scattering amplitude on the right, we will obtain zero by the Ward identity, and the contraction of $\epsilon_+^{*\nu}$ with the electron emission vertex similarly gives zero. Thus, for the purpose of computing the singular term as the photon momentum q goes on-shell, we may replace

$$\frac{-ig^{\mu\nu}}{q^2} \rightarrow \frac{+i}{q^2} \sum_i \epsilon_{Ti}^\mu \epsilon_{Ti}^{\nu *} \tag{17.81}$$

and evaluate the photon emission and absorption amplitudes with transverse polarization vectors.

Matrix Element for Electron Splitting

By replacing the numerator of the intermediate propagator with a sum over polarization vectors, we decouple the photon or electron emission vertex from the rest of the diagram. We will now evaluate this vertex explicitly between physical polarization states of massless particles. The kinematics is shown in Fig. 17.16. The two final particles should be almost collinear, with a small relative transverse momentum. We can choose the incident electron momentum to lie along the $\hat{3}$ axis and the outgoing momenta to lie in the $\hat{1}$-$\hat{3}$ plane. Let z be the fraction of the energy of the initial electron that is carried off by the photon. Then the three 4-momenta can be written as

$$\begin{aligned} p &= (p, 0, 0, p), \\ q &\approx (zp, p_\perp, 0, zp), \\ k &\approx ((1-z)p, -p_\perp, 0, (1-z)p). \end{aligned} \tag{17.82}$$

These three vectors satisfy $p^2 = q^2 = k^2 = 0$, up to terms of order $p_\perp^2$.

In the process where a real photon is emitted, we should have p^2 and q^2 exactly zero, and k^2 slightly off-shell by an amount of order $p_\perp^2$. We will need to know the value of k^2, which appears in the virtual electron propagator. So let us modify Eqs. (17.82) to satisfy the condition $q^2 = 0$ up to terms of order

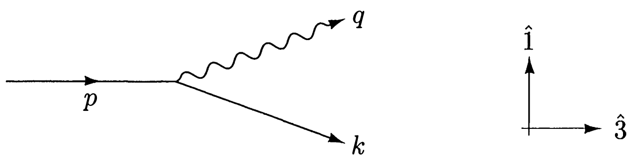


Figure 17.16. Kinematics of the vertex for emission of a collinear electron or photon.

$p_\perp^4$, rewriting q and k as

$$\begin{aligned} q &= (zp, p_\perp, 0, zp - \frac{p_\perp^2}{2zp}), \\ k &= ((1-z)p, -p_\perp, 0, (1-z)p + \frac{p_\perp^2}{2zp}). \end{aligned} \tag{17.83}$$

With this modification,

$$k^2 = -p_\perp^2 - 2(1-z)\frac{p_\perp^2}{2z} + \mathcal{O}(p_\perp^4).$$

Thus, if the photon is real and the electron is virtual, we have

$$q^2 = 0, \qquad k^2 = -\frac{p_\perp^2}{z}. \tag{17.84}$$

Reciprocally, in the process with a real electron and a virtual photon,

$$k^2 = 0, \qquad q^2 = -\frac{p_\perp^2}{(1-z)}. \tag{17.85}$$

These more accurate expressions will be needed only in the propagator of the virtual particle. The matrix element of the electron-photon vertex begins in order $p_\perp$, so it is not significantly affected by the modification of (17.82) to (17.83), and is the same (to lowest order) no matter which particle is virtual.

We now calculate the matrix elements of the QED vertex between massless states of definite helicity. If the initial electron is left-handed, the final electron must also be left-handed, by helicity conservation. Then the photon emission vertex is given by

$$i\mathcal{M} = \bar{u}_L(k)(-ie\gamma_\mu)u_L(p)\epsilon_T^{*\mu}(q), \tag{17.86}$$

where the photon polarization vector may be either left- or right-handed. Recalling the helicity-basis expressions

$$\gamma_\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \bar{\sigma}_\mu & 0 \end{pmatrix}, \qquad u_L(p) = \sqrt{2p^0}\begin{pmatrix} \xi(p) \\ 0 \end{pmatrix} \quad (\text{for } m = 0),$$

we can write more explicitly

$$i\mathcal{M} = ie\sqrt{2(1-z)p}\sqrt{2p}\,\xi^\dagger(k)\sigma^i\xi(p)\,\epsilon_T^{*i}(q). \tag{17.87}$$

To order $p_\perp$, the left-handed spinors are

$$\xi(p) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \xi(k) = \begin{pmatrix} p_\perp/2(1-z)p \\ 1 \end{pmatrix}. \tag{17.88}$$

The polarization vectors for the photon are

$$\epsilon_L^{*i}(q) = \frac{1}{\sqrt{2}}\left(1, i, -\frac{p_\perp}{zp}\right), \qquad \epsilon_R^{*i}(q) = \frac{1}{\sqrt{2}}\left(1, -i, -\frac{p_\perp}{zp}\right). \tag{17.89}$$

Notice that, when these vectors are contracted with the Pauli matrix in Eq. (17.87), the first two components of the right-handed polarization vector give $(\sigma^1 - i\sigma^2) = 2\sigma^-$, which annihilates $\xi(p)$. The only remaining term comes from the $i = 3$ component, and we find

$$i\mathcal{M}(e_L^- \to e_L^- \gamma_R) = ie\frac{\sqrt{2(1-z)}}{z}p_\perp. \tag{17.90}$$

For the left-handed photon polarization, there is an additional contribution from the first two components of ϵ_L^*. These add to

$$i\mathcal{M}(e_L^- \to e_L^- \gamma_L) = ie\frac{\sqrt{2(1-z)}}{z(1-z)}p_\perp. \tag{17.91}$$

Parity invariance implies that the values of the matrix elements are unchanged if all helicities are flipped; this immediately gives the required matrix elements for the case of an initial e_R^-. The squared matrix element, averaged over initial helicities, is therefore

$$\frac{1}{2}\sum_{\text{pols.}} |\mathcal{M}|^2 = \frac{2e^2 p_\perp^2}{z(1-z)}\left[\frac{1 + (1-z)^2}{z}\right]. \tag{17.92}$$

The first term in the brackets comes from a photon with spin parallel to the electron spin; the second term comes from a photon with spin opposite to the electron spin.

The Equivalent Photon Approximation

Now we have all the pieces needed to compute the cross sections for the processes shown in Fig. 17.15. We first consider the process with a virtual photon. Call the initial state on the right-hand side of the diagram X and the final state Y, and let $\mathcal{M}_{\gamma X}$ represent the matrix element for the scattering of the photon from X. We will assume for simplicity that X is unpolarized, so that the scattering cross section does not depend on the virtual photon polarization. Then the complete diagram gives a cross section

$$\sigma = \frac{1}{(1+v_X)2p2E_X}\int \frac{d^3k}{(2\pi)^3}\frac{1}{2k^0}\int d\Pi_Y\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\left(\frac{1}{q^2}\right)^2|\mathcal{M}_{\gamma X}|^2, \tag{17.93}$$

where v_X is the velocity of X and $\int d\Pi_Y$ is the phase space integral over Y.

The integral has a singularity when k is collinear with the incident electron momentum p. To isolate the singularity, substitute for k^0 and q^2 from Eqs.

(17.82) and (17.85) and rewrite the integral over k as

$$d^3k = dk^3 d^2k_\perp = pdz \cdot \pi dp_\perp^2. \tag{17.94}$$

Then the cross section can be expressed as

$$\begin{aligned}\sigma &= \int \frac{pdzdp_\perp^2}{16\pi^2(1-z)p}\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\frac{(1-z)^2}{p_\perp^4}\frac{z}{(1+v_X)2zp2E_X}\int d\Pi_Y|\mathcal{M}_{\gamma X}|^2\\ &= \int \frac{dzdp_\perp^2}{16\pi^2(1-z)}\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\frac{z(1-z)^2}{p_\perp^4}\cdot\sigma(\gamma X\to Y).\end{aligned} \tag{17.95}$$

Finally, insert the spin-averaged electron emission vertex (17.92), to obtain

$$\begin{aligned}\sigma &= \int \frac{dzdp_\perp^2}{16\pi^2}\frac{z(1-z)}{p_\perp^4}\frac{2e^2p_\perp^2}{z(1-z)}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(\gamma X\to Y)\\ &= \int_0^1 dz\int\frac{dp_\perp^2}{p_\perp^2}\frac{\alpha}{2\pi}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(\gamma X\to Y).\end{aligned} \tag{17.96}$$

The integral over $p_\perp^2$ runs from momentum transfers of order s down to the electron mass m^2, which cuts off the singularity. Thus, our final result is

$$\sigma(e^-X\to e^-Y) = \int_0^1 dz\,\frac{\alpha}{2\pi}\log\frac{s}{m^2}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(\gamma X\to Y). \tag{17.97}$$

The cross section on the right-hand side is computed for a real, transversely polarized photon of momentum zp. The factor $\log(s/m^2)$ represents the mass singularity. This formula is the Weizsäcker-Williams *equivalent photon approximation*, which we encountered earlier in Problems 5.5 and 6.2.

Formula (17.97) takes on a new significance when it is juxtaposed with the QCD predictions of the previous two sections. This QED formula has just the same form as a parton model expression, with the Weizsäcker-Williams distribution function

$$f_\gamma(z) = \frac{\alpha}{2\pi}\log\frac{s}{m^2}\left[\frac{1+(1-z)^2}{z}\right] \tag{17.98}$$

playing the role of the probability to find a photon of longitudinal fraction z in the incident electron.

The Electron Distribution

The first diagram of Fig. 17.15, with an emitted photon and a virtual electron, can be treated in the same way. The analogue of (17.93) is

$$\sigma = \frac{1}{(1+v_X)2p2E_X}\int\frac{d^3q}{(2\pi)^3}\frac{1}{2q^0}\int d\Pi_Y\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\left(\frac{1}{k^2}\right)^2|\mathcal{M}_{e^-X}|^2.$$

Following the steps that led to (17.97), we find

$$\begin{aligned}\sigma(e^-X \to \gamma Y) &= \int \frac{dz dp_\perp^2}{16\pi^2 z}\left[\frac{1}{2}\sum|\mathcal{M}|^2\right]\frac{z^2}{p_\perp^4}\cdot(1-z)\sigma(e^-X\to Y)\\ &= \int \frac{dz dp_\perp^2}{16\pi^2}\frac{z(1-z)}{p_\perp^4}\frac{2e^2p_\perp^2}{z(1-z)}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(e^-X\to Y)\\ &= \int_0^1 dz\int\frac{dp_\perp^2}{p_\perp^2}\frac{\alpha}{2\pi}\left[\frac{1+(1-z)^2}{z}\right]\cdot\sigma(e^-X\to Y), \qquad (17.99)\end{aligned}$$

where the intermediate electron carries a longitudinal fraction $(1-z)$.

It is tempting to substitute $x = (1-z)$ and interpret the factor multiplying the cross section under the integral in (17.99) as the parton distribution for finding an electron parton in the electron. This would give

$$f_e^{(1)}(x) = \frac{\alpha}{2\pi}\log\frac{s}{m^2}\left[\frac{1+x^2}{1-x}\right]. \qquad (17.100)$$

However, this expression is not adequate. Most obviously, it does not take into account the processes without radiation, in which the electron remains an electron at the full energy. This is easily remedied by considering (17.100) as the order-α correction to the most naive expectation,

$$f_e^{(0)}(x) = \delta(1-x), \qquad (17.101)$$

in which we consider the electron to contain only an electron at the full energy. Unfortunately, the sum of (17.101) and (17.100) still does not give an adequate description of the electron distribution, for two reasons. First, Eq. (17.100) diverges near $x = 1$, and we need a prescription for treating this singularity. Second, while Eq. (17.100) takes into account the virtual electrons moved to longitudinal fraction x from $x = 1$ by radiation, it does not take into account the concomitant loss of electrons from the delta function peak at $x = 1$.

The divergence of (17.100) at $x = 1$ corresponds to the emission of soft photons. We saw in Section 6.5 that the emission of soft photons does not affect the rate of a QED reaction. Order by order in α, one finds that infrared-divergent positive contributions to the total rate from the emission of soft photons are balanced by negative contributions from diagrams with soft virtual photons. In the present example, the negative contribution must decrease the weight of the process in which no photon is emitted. Thus, to order α, the parton distribution for electrons in the electron should have the form

$$f_e(x) = \delta(1-x) + \frac{\alpha}{2\pi}\log\frac{s}{m^2}\left(\frac{1+x^2}{(1-x)} - A\delta(1-x)\right). \qquad (17.102)$$

The coefficient A results from diagrams with virtual photons that we have not computed. However, the effect of these diagrams is easy to understand; they subtract from the delta function the probability that has been moved

to lower x by radiation, so that the integral over the full term of order α is zero. Another way of expressing this criterion is that A is determined by the condition that the electron contain exactly one electron parton,

$$\int_0^1 dx\, f_e(x) = 1. \tag{17.103}$$

(This equation will be modified below, when we include pair-creation processes.)

It is not so clear how to integrate over the singular denominator in (17.100) to determine A explicitly. It is conventional to define a distribution that can be integrated by subtracting a delta function from the singular term. Define the distribution

$$\frac{1}{(1-x)_+} \tag{17.104}$$

to agree with the function $1/(1-x)$ for all values of x less than 1, and to have a singularity at $x = 1$ such that the integral of this distribution with any smooth function $f(x)$ gives

$$\int_0^1 dx\, \frac{f(x)}{(1-x)_+} = \int_0^1 dx\, \frac{f(x) - f(1)}{(1-x)}. \tag{17.105}$$

Less formally,

$$\frac{1}{(1-x)_+} = \lim_{\epsilon \to 0} \left[\frac{1}{(1-x)} \theta(1 - x - \epsilon) - \delta(1 - x) \int_0^{1-\epsilon} dx' \frac{1}{(1-x')} \right]. \tag{17.106}$$

The more formal definition (17.105) is often easier to use in practice.

Using this definition, we can bring a piece of the delta function into the singular term of (17.102) by changing the denominator $(1-x)$ to $(1-x)_+$. Then, to normalize (17.102), we need the integral

$$\int_0^1 dx\, \frac{1 + x^2}{(1-x)_+} = \int_0^1 dx\, \frac{x^2 - 1}{(1-x)} = -\frac{3}{2}.$$

Our final form of the electron distribution, to order α, is

$$f_e(x) = \delta(1 - x) + \frac{\alpha}{2\pi} \log \frac{s}{m^2} \left[\frac{1 + x^2}{(1-x)_+} + \frac{3}{2} \delta(1 - x) \right]. \tag{17.107}$$

This distribution is now properly normalized, but it is still highly singular near $x = 1$. Thus, we should expect higher-order corrections to the electron distribution function to be important in this region. We must, then, think about how to treat the emission of many collinear photons.

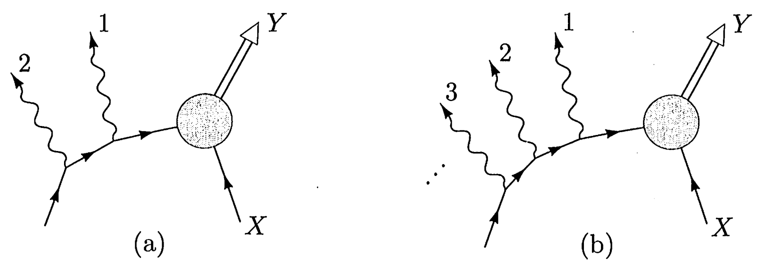


Figure 17.17. Higher-order diagrams with collinear photon emission: (a) two collinear photons; (b) many collinear photons.

Multiple Splittings

In fact, it is not difficult to extend the analysis we have just completed to account for emission of many collinear photons. Consider the process shown in Fig. 17.17(a), in which photon 1 is radiated with a transverse momentum $p_{1\perp}$ and photon 2 with a transverse momentum $p_{2\perp}$. The emission of photon 2 can be computed just as we did above. If $p_{2\perp} \ll p_{1\perp}$, the first virtual electron is very close to mass shell compared to $p_{1\perp}^2$ and so we can ignore its virtuality in computing the emission of the photon 1. The double photon emission gives a contribution of order

$$\left(\frac{\alpha}{2\pi}\right)^2 \int_{m^2}^{s} \frac{dp_{1\perp}^2}{p_{1\perp}^2} \int_{m^2}^{p_{1\perp}^2} \frac{dp_{2\perp}^2}{p_{2\perp}^2} = \frac{1}{2}\left(\frac{\alpha}{2\pi}\right)^2 \log^2 \frac{s}{m^2}.$$

In the opposite limit, $p_{2\perp} \gg p_{1\perp}$, there is no denominator of order $p_{1\perp}^2$, and so we do not find a double logarithm. Only in the case $p_{2\perp} \ll p_{1\perp}$ can the contribution of order α^2 compete with the contribution of order α.

This argument extends to the emission of arbitrarily many collinear photons, Fig. 17.17(b). The region of integration over the photon phase space corresponding to the ordering

$$p_{1\perp} \gg p_{2\perp} \gg p_{3\perp} \gg \cdots \tag{17.108}$$

gives a contribution that contains the factor

$$\frac{1}{n!}\left(\frac{\alpha}{2\pi}\right)^n \log^n \frac{s}{m^2}. \tag{17.109}$$

If the photon transverse momenta are ordered in any other way, the contribution from that region contains at least one less power of the large logarithm at the same order in α. If condition (17.108) holds, the virtual electron momenta are increasingly off-mass-shell as one proceeds from the outside of the diagram toward the hard collision. In this case, the electron momenta are said to be *strongly ordered.*

This set of conclusions has an interesting physical interpretation. Since the intermediate electrons are increasingly virtual as we go into the diagram, it is natural to interpret them as components of the physical electron when this particle is analyzed at successively smaller distance scales. The intermediate electron with $k^2 \sim p_\perp^2$ may be thought of as a constituent of the electron made visible when the wavefunction of the physical electron is probed with a resolution $\Delta r \sim (p_\perp)^{-1}$. In this picture, the electron seen at one resolution can be resolved at a finer scale into a more virtual electron and a number of photons.

From either the perspective of computing Feynman diagrams or the grander perspective of the electron structure, it is useful to imagine the splitting of the electron into a virtual electron plus photons to be a continuous evolution process as a function of the transverse momentum of the electron constituent. To describe this process mathematically, we introduce an explicit $p_\perp$ dependence of the electron and photon distribution functions. We define the functions $f_\gamma(x, Q)$ and $f_e(x, Q)$ to give the probabilities of finding a photon or an electron of longitudinal fraction x in the physical electron, taking into account collinear photon emissions with transverse momenta $p_\perp < Q$. If Q is slightly increased to $Q + \Delta Q$, we must take into account the possibility that an electron constituent in $f_e(x, Q)$ will radiate a photon with $Q < p_\perp < Q + \Delta Q$. The differential probability for an electron to split off a photon that carries away a fraction z of its energy is

$$\frac{\alpha}{2\pi} \frac{dp_\perp^2}{p_\perp^2} \frac{1 + (1-z)^2}{z}. \tag{17.110}$$

The new photon distribution can therefore be computed as follows:

$$\begin{aligned} f_\gamma(x, Q+\Delta Q) &\\ = f_\gamma(x, Q) &+ \int_0^1 dx' \int_0^1 dz \left[\frac{\alpha}{2\pi} \frac{\Delta Q^2}{Q^2} \frac{1 + (1-z)^2}{z}\right] f_e(x', p_\perp) \delta(x - zx') \\ = f_\gamma(x, Q) &+ \frac{\Delta Q}{Q} \int_x^1 \frac{dz}{z} \left[\frac{\alpha}{\pi} \frac{1 + (1-z)^2}{z}\right] f_e(\frac{x}{z}, p_\perp). \end{aligned} \tag{17.111}$$

Passing to a continuous evolution, we find that the function $f_\gamma(x, Q)$ is determined by the integral-differential equation

$$\frac{d}{d \log Q} f_\gamma(x, Q) = \int_x^1 \frac{dz}{z} \left[\frac{\alpha}{\pi} \frac{1 + (1-z)^2}{z}\right] f_e(\frac{x}{z}, Q). \tag{17.112}$$

Similarly, the distribution of component electrons in the physical electron will evolve with Q, reflecting the appearance of electrons at lower values of x due to photon radiation, and the disappearance of these electrons at higher x. The term in brackets in Eq. (17.107) gives a correct accounting of both effects

for the radiation of a single photon. Thus, the electron distribution evolves according to

$$\frac{d}{d\log Q} f_e(x,Q) = \int_x^1 \frac{dz}{z}\left[\frac{\alpha}{\pi}\left(\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z)\right)\right] f_e\left(\frac{x}{z},Q\right). \qquad (17.113)$$

By integrating these integral-differential equations using appropriate initial conditions, we sum all of the logarithmically enhanced terms of the form (17.109). The initial conditions should be fixed at a point that will reproduce the correct denominator of the logarithms in Eqs. (17.98) and (17.107). Thus, we should set

$$f_e(x,Q) = \delta(1-x), \qquad f_\gamma(x,Q) = 0, \qquad (17.114)$$

at $Q^2 \sim m^2$.

The resulting distribution functions can be used to compute the cross sections for electron hard scattering from an arbitrary target. Then Eqs. (17.97) and (17.99) should be replaced by

$$\begin{aligned} \sigma(e^- X \to e^- + n\gamma + Y) &= \int_0^1 dx\, f_\gamma(x,Q)\sigma(\gamma X \to Y), \\ \sigma(e^- X \to n\gamma + Y) &= \int_0^1 dx\, f_e(x,Q)\sigma(e^- X \to Y), \end{aligned} \qquad (17.115)$$

where the cross sections under the integrals are computed for a photon or electron carrying a fraction x of the original electron momentum, the functions $f_\gamma(x,Q)$, $f_e(x,Q)$ are the solutions to Eqs. (17.112) and (17.113), and the momentum Q is chosen as a characteristic momentum transfer of the γX or $e^- X$ subprocess.

Photon Splitting to Pairs

The evolution equations for $f_\gamma(x)$ and $f_e(x)$ need one more modification before they can be considered complete. As written, these equations account for the radiation of photons by electrons to all orders. However, they omit another process that is of the same order in α: the splitting of a photon into an electron-positron pair. We must include this process in our evolution equations, because the process shown in Fig. 17.18, for example, has the same logarithmic enhancement as that shown in Fig. 17.17(a).

We can compute the effects of photon splitting in the same way that we computed with effects of photon radiation. The basic kinematics of the process is very similar, as shown in Fig. 17.19; the only difference is that the photon is now in the initial state, while the final state consists of an almost collinear electron-positron pair. We need to work out the analogue of Eq. (17.92) for this process.

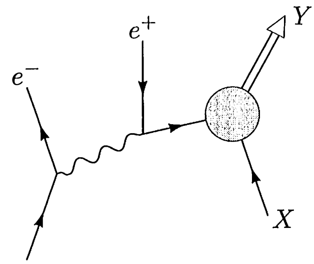


Figure 17.18. A process that involves e^+e^- pair creation enhanced by a collinear mass singularity.

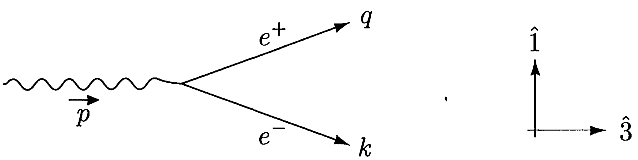


Figure 17.19. Kinematics of the vertex for photon conversion to a collinear electron-positron pair.

Consider the case in which the outgoing electron is left-handed. Then the outgoing positron must be right-handed, by helicity conservation; its spin wavefunction will contain a left-handed spinor. Let us take the electron momentum to be k, given by Eq. (17.82), and the positron momentum to be q. Then the vertex gives a matrix element

$$i\mathcal{M} = -ie\bar{u}_L(k)\gamma_\mu v_L(q)\epsilon^\mu_T(p), \tag{17.116}$$

where the photon polarization vector can be either left- or right-handed. When we insert the explicit forms for the massless spinors, we obtain

$$i\mathcal{M} = ie\sqrt{2(1-z)p}\sqrt{2zp}\,\xi^\dagger(k)\sigma^i\xi(q)\cdot\epsilon^i_T(p),$$

where the electron and positron spinors are given, to order $p_\perp$, by

$$\xi(q) = \begin{pmatrix} -p_\perp/2zp \\ 1 \end{pmatrix}, \qquad \xi(k) = \begin{pmatrix} p_\perp/2(1-z)p \\ 1 \end{pmatrix}.$$

The polarization vectors for the photon are

$$\epsilon^i_L(p) = \frac{1}{\sqrt{2}}(1,-i,0), \qquad \epsilon^i_R(p) = \frac{1}{\sqrt{2}}(1,i,0).$$

Dotting these vectors into σ^i, we find for the polarized matrix elements

$$i\mathcal{M}(\gamma_L \to e^-_L e^+_R) = -ie\frac{\sqrt{2z(1-z)}}{z}p_\perp,$$

and

$$i\mathcal{M}(\gamma_R \to e_L^- e_R^+) = +ie\frac{\sqrt{2z(1-z)}}{(1-z)}p_\perp.$$

Again, the matrix elements are unchanged if all helicities are flipped. Thus the squared matrix element, averaged over initial photon polarizations, is

$$\frac{1}{2}\sum_{\text{pols.}}|\mathcal{M}|^2 = \frac{2e^2p_\perp^2}{z(1-z)}\left[z^2+(1-z)^2\right], \tag{17.117}$$

where z is the momentum fraction carried by the positron. The first term in the brackets comes from processes in which the spin of the positron is parallel to the spin of the photon; the second term comes from processes where the electron spin is parallel to the photon spin.

The squared matrix element (17.117) generates an evolution of constituent photons into electrons and positrons. The form of the evolution equation is similar to (17.113), but with the photon distribution on the right-hand side, and with the expression in parentheses replaced by

$$\left(z^2+(1-z)^2\right). \tag{17.118}$$

When we create an electron-positron pair, we must remove a photon; this requires a negative term in the evolution equation for the photon distribution (17.112) that contains a delta function multiplying the normalization of (17.118):

$$\int_0^1 dz\left(z^2+(1-z)^2\right) = \frac{2}{3}. \tag{17.119}$$

Evolution Equations for QED

Including the effects of pair creation, we find the complete evolution equations for electron, positron, and photon distributions in QED. These equations, originally derived by Gribov and Lipatov, sum the leading logarithms from collinear singularities to all orders in α. The evolution equations take the form

$$\begin{aligned}
\frac{d}{d\log Q}f_\gamma(x,Q) &= \frac{\alpha}{\pi}\int_x^1\frac{dz}{z}\Big\{P_{\gamma\leftarrow e}(z)\left[f_e(\tfrac{x}{z},Q)+f_{\bar e}(\tfrac{x}{z},Q)\right] \\
&\qquad\qquad + P_{\gamma\leftarrow\gamma}(z)f_\gamma(\tfrac{x}{z},Q)\Big\}, \\
\frac{d}{d\log Q}f_e(x,Q) &= \frac{\alpha}{\pi}\int_x^1\frac{dz}{z}\Big\{P_{e\leftarrow e}(z)f_e(\tfrac{x}{z},Q)+P_{e\leftarrow\gamma}(z)f_\gamma(\tfrac{x}{z},Q)\Big\}, \\
\frac{d}{d\log Q}f_{\bar e}(x,Q) &= \frac{\alpha}{\pi}\int_x^1\frac{dz}{z}\Big\{P_{e\leftarrow e}(z)f_{\bar e}(\tfrac{x}{z},Q)+P_{e\leftarrow\gamma}(z)f_\gamma(\tfrac{x}{z},Q)\Big\}.
\end{aligned} \tag{17.120}$$

The *splitting functions* $P_{i\leftarrow j}(z)$ are given by

$$\begin{aligned}
P_{e\leftarrow e}(z) &= \frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z), \\
P_{\gamma\leftarrow e}(z) &= \frac{1+(1-z)^2}{z}, \\
P_{e\leftarrow\gamma}(z) &= z^2 + (1-z)^2, \\
P_{\gamma\leftarrow\gamma}(z) &= -\frac{2}{3}\delta(1-z).
\end{aligned} \tag{17.121}$$

To obtain the distribution functions for an electron relevant to a given momentum transfer Q, we should integrate these equations with the initial conditions

$$f_e(x,Q) = \delta(1-x), \qquad f_{\bar{e}}(x,Q) = 0, \qquad f_\gamma(x,Q) = 0, \tag{17.122}$$

at $Q = m$. With different initial conditions, the same equations give the distribution functions for a physical positron or photon. The solutions to these equations are used as in Eq. (17.115) to compute cross sections involving processes induced by electrons, positrons, or photons that involve large momentum transfer.

The evolution equations (17.120) are constructed in such a way as to conserve electron number and longitudinal momentum. Thus, the basic sum rules (17.36) and (17.39) satisfied by the parton distributions of hadrons also apply to the QED distribution functions. Specifically, the distribution functions of the electron contain one net electron constituent,

$$\int_0^1 dx\,\big[f_e(x,Q) - f_{\bar{e}}(x,Q)\big] = 1, \tag{17.123}$$

and account for the total momentum of the physical electron,

$$\int_0^1 dx\, x\big[f_e(x,Q) + f_{\bar{e}}(x,Q) + f_\gamma(x,Q)\big] = 1. \tag{17.124}$$

It is an instructive exercise to verify explicitly, using Eq. (17.120), that the values of these integrals do not depend on Q.

The Altarelli-Parisi Equations

If we encounter mass singularities in QED associated with collinear photon emission, we must also encounter mass singularities in QCD associated with collinear gluon and quark emission. If we compute the corrections of order α_s to the leading-order parton cross sections discussed in Sections 17.3 and 17.4, using massless quarks and gluons, we will find that these correction terms

diverge when we integrate over the collinear configurations. Thus the parton-model expressions, at least in their simplest form, break down already at the next-to-leading order in α_s.

However, assuming that the singularities of QCD are no worse than those of QED, the considerations of the previous section tell us how to treat these singular terms. In QED, we found it natural to include the large corrections associated with the mass singularities in the parton distributions rather than in the hard-scattering cross sections. Viewed in this way, the singular terms supply the kernel of an evolution equation for the parton distributions as a function of the logarithm of the momentum scale. Hard scattering with a momentum transfer Q probes the electron at a distance of order Q^{-1}. When the electron wavefunction is resolved to very small scales, it appears as a constituent electron, carrying only a fraction of the total longitudinal momentum, plus a number of constituent photons and electron-positron pairs. Any one of these constituents that carries a substantial fraction of the total electron momentum can initiate a hard-scattering process.

Precisely the same logic applies to the calculation of QCD cross sections. The contributions from the region of collinear gluon or quark emission should be associated with the parton distribution functions rather than with the hard-scattering cross sections. If we make this association, we find that the parton distributions are no longer independent of the momentum Q that characterizes the hard-scattering process; rather, they now evolve logarithmically with Q. For example, the basic equation (17.30) for deep-inelastic scattering will become

$$\frac{d^2\sigma}{dxdy}(e^-p \to e^-X) = \Bigl(\sum_f xf_f(x,Q)Q_f^2\Bigr) \cdot \frac{2\pi\alpha^2 s}{Q^4}\bigl[1 + (1-y)^2\bigr], \qquad (17.125)$$

and so Bjorken scaling will be violated. Since this violation takes place only on a logarithmic scale in Q^2, it will be a subtle effect, and *approximate* Bjorken scaling will still be a prediction of QCD. But the violation of Bjorken scaling is inevitable, since QCD is a quantum field theory with degrees of freedom at all momentum scales. As we probe the proton wavefunction at increasingly short distances, we excite the high-momentum degrees of freedom and resolve the wavefunction into an increasing number of quarks, antiquarks, and gluons.

The evolution of the QED parton distributions, governed by Eq. (17.120), is characterized by the parameter α/π, so the parton distributions change by $\sim$1% as Q is changed by a factor of 10. In QCD, the corresponding factor governing the rate of evolution should be $\alpha_s(Q)/\pi$. Thus, when Q is very small, the evolution is rapid and contributions of higher order in perturbation theory are important. Ultimately, the initial conditions for the evolution are determined by the form of the proton wavefunction at large distance scales, which cannot be calculated using Feynman diagrams. On the other hand, when Q is large, well above 1 GeV in practice, the evolution becomes slow and is dominated by the leading order in perturbation theory. In that case,

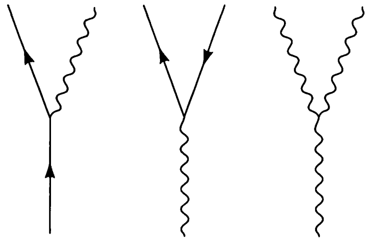


Figure 17.20. The three vertices that contribute to parton evolution in QCD.

QCD perturbation theory makes precise predictions for the form of the evolution of the parton distributions, and these predictions can be tested against experiment.

To derive the evolution equations of parton distributions in QCD we can use the same techniques and logic that we used above for QED. There is a subtlety, that the reduction of the gluon propagator to transverse polarization states in the limit $q^2 \to 0$, Eq. (17.81), cannot be proved so simply as in QED. However, the result is correct also in the non-Abelian case.* Once this technical point is resolved, the kinematics of collinear emission is exactly the same as in QED. Thus we find evolution equations of the same form as in QED, modified only by the replacement of α by α_s, the insertion of appropriate color factors, and the accounting of the effects of the three-gluon vertex.

Collinear emission processes in QCD involve the three vertices shown in Fig. 17.20. Of these, the first two have the same Lorentz structure as those shown in Figs. 17.16 and 17.19. The only difference, aside from the strength of the coupling constant, comes in the color indices. We will treat color just as we treated spin in the preceding analysis: We average over initial colors, and sum over final colors. Then the first vertex of Fig. 17.20, representing the splitting of a quark into a quark and a gluon, receives the color factor

$$\tfrac{1}{3}\operatorname{tr}[t^a t^a] = C_2(r) = \tfrac{4}{3}. \tag{17.126}$$

The second vertex, representing the splitting of a gluon into a quark-antiquark pair, receives the factor

$$\tfrac{1}{8}\operatorname{tr}[t^a t^a] = \tfrac{1}{2}. \tag{17.127}$$

The third vertex in Fig. 17.20 represents the splitting of a gluon to two gluons, an effect that is new to the non-Abelian case. It is straightforward to compute the contribution of this vertex to the evolution equations by taking the matrix elements of the vertex between transverse gluon states of definite helicity. This calculation is the subject of Problem 17.4.

*See, for example, J. Collins and D. Soper, in A. Mueller, *Quantum Chromodynamics* (World Scientific, Singapore, 1991).

By accounting for all of these effects, we can modify the QED evolution equations (17.120) into the correct set of evolution equations for parton distributions in QCD. These are known as the *Altarelli-Parisi equations.* They describe the coupled evolution of parton distributions $f_f(x,Q)$, $f_{\bar{f}}(x,Q)$ for each flavor of quark and antiquark that can be treated as massless at the scale Q, together with the parton distribution of gluons, $f_g(x,Q)$. Explicitly,

$$\begin{aligned}
\frac{d}{d\log Q} f_g(x,Q) &= \frac{\alpha_s(Q^2)}{\pi} \int_x^1 \frac{dz}{z} \Big\{ P_{g\leftarrow q}(z) \sum_f \big[f_f(\tfrac{x}{z},Q) + f_{\bar{f}}(\tfrac{x}{z},Q) \big] \\
&\qquad\qquad + P_{g\leftarrow g}(z) f_g(\tfrac{x}{z},Q) \Big\}, \\
\frac{d}{d\log Q} f_f(x,Q) &= \frac{\alpha_s(Q^2)}{\pi} \int_x^1 \frac{dz}{z} \Big\{ P_{q\leftarrow q}(z) f_f(\tfrac{x}{z},Q) + P_{q\leftarrow g}(z) f_g(\tfrac{x}{z},Q) \Big\}, \\
\frac{d}{d\log Q} f_{\bar{f}}(x,Q) &= \frac{\alpha_s(Q^2)}{\pi} \int_x^1 \frac{dz}{z} \Big\{ P_{q\leftarrow q}(z) f_{\bar{f}}(\tfrac{x}{z},Q) + P_{q\leftarrow g}(z) f_g(\tfrac{x}{z},Q) \Big\}.
\end{aligned} \tag{17.128}$$

The first three splitting functions can be taken from Eqs. (17.121), multiplied by the color factors computed in Eqs. (17.126) and (17.127):

$$\begin{aligned}
P_{q\leftarrow q}(z) &= \frac{4}{3}\left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z) \right], \\
P_{g\leftarrow q}(z) &= \frac{4}{3}\left[\frac{1+(1-z)^2}{z} \right], \\
P_{q\leftarrow g}(z) &= \frac{1}{2}\left[z^2 + (1-z)^2 \right].
\end{aligned} \tag{17.129}$$

The fourth splitting function requires also the computation of Problem 17.4; the result is

$$P_{g\leftarrow g}(z) = 6\left[\frac{(1-z)}{z} + \frac{z}{(1-z)_+} + z(1-z) + \Big(\frac{11}{12} - \frac{n_f}{18}\Big)\delta(1-z) \right]. \tag{17.130}$$

The final term in this expression, which is proportional to n_f, the number of light quark flavors, is the subtraction term associated with gluon splitting into $q\bar{q}$ pairs. The Altarelli-Parisi equations describe the evolution of parton distributions for any hadron, or any hadronic constituent, up to corrections of order α_s that are not enhanced by large logarithms.

Our derivation of the Altarelli-Parisi equations respects the conservation laws of QCD for quark numbers and longitudinal momentum. Thus, the equations must respect the parton-model sum rules (17.36) and (17.39). As in the QED case, it is instructive to verify explicitly that these integrals are independent of Q.

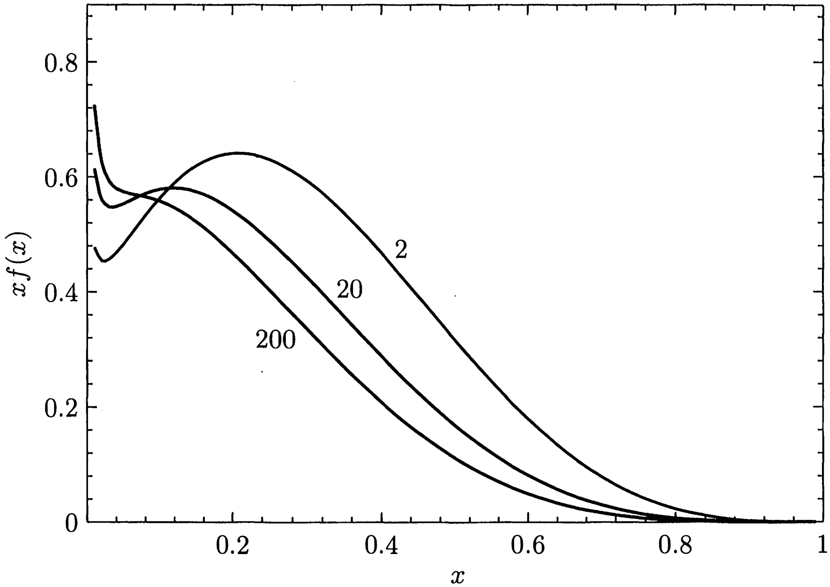


Figure 17.21. The u quark parton distribution function $xf_u(x, Q)$ at $Q =$ 2, 20, and 200 GeV, showing the effects of parton evolution according the Altarelli-Parisi equations. These curves are taken from the CTEQ fit to deep inelastic scattering data described in Fig. 17.6.

In QED, we could use the evolution equations to explicitly compute the structure function of the electron. In QCD, this is no longer possible, because the initial conditions required to integrate the equations are determined by the strong-coupling region of QCD and so are not known *a priori*. However, one can determine the initial conditions of the proton structure experimentally, by measuring the cross section for deep inelastic scattering at a given value of Q^2. One can then predict the structure functions, and thus the deep inelastic cross sections, at higher values of Q^2. There is one subtlety in this analysis: The gluon distribution is not directly measured in deep inelastic scattering, but it does enter the evolution equation for the quark distributions. Thus, some of the information on the Q^2 dependence of deep inelastic scattering simply goes into determining the gluon distribution. However, the gluon distribution is absolutely normalized by the momentum sum rule (17.39), so the evolution equations have predictive power even if this distribution must be fit from the data.

The Altarelli-Parisi equations predict a characteristic form for the evolution of parton distribution functions, shown in Fig. 17.21. Partons at high x tend to radiate and drop down to lower values of x. Meanwhile, new partons are formed at low x as products of this radiation. Thus, the parton distributions decrease at large x and increase much more rapidly at small x as Q^2 increases. We can picture the proton as having more and more constituents,

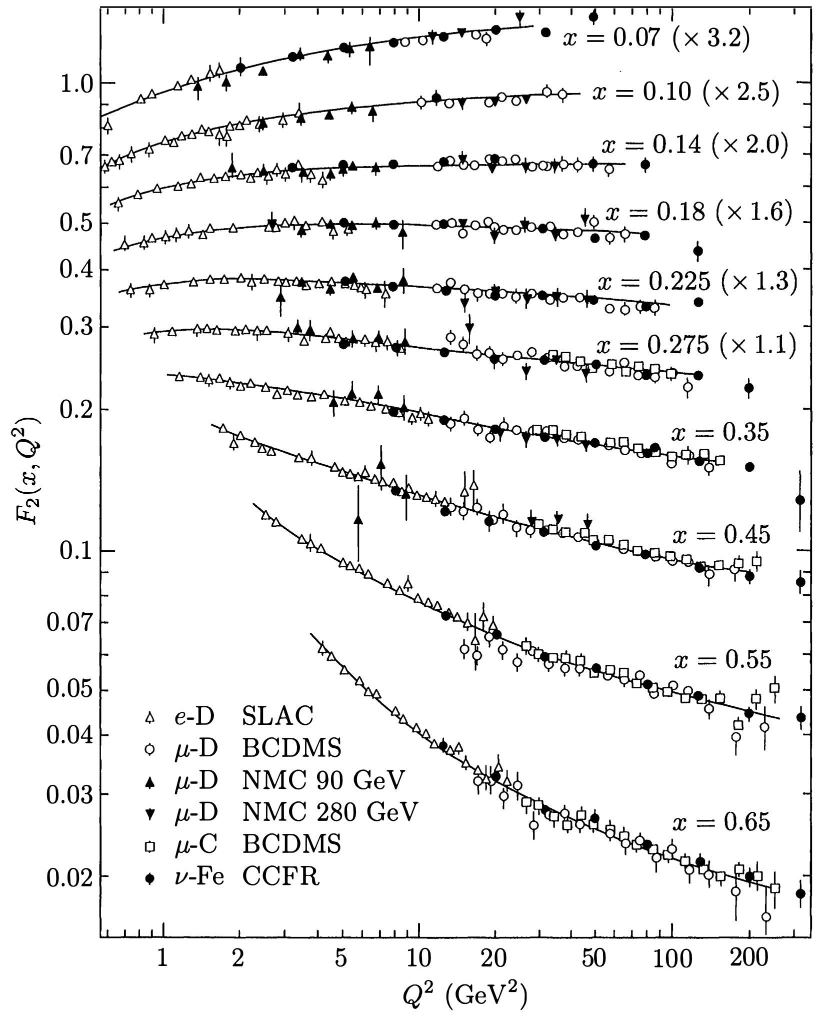


Figure 17.22. Dependence on Q^2 of the combination of quark distribution functions $F_2 = \sum_f xQ_f^2 f_f(x, Q^2)$ measured in deep inelastic electron-proton scattering. The various curves show the variation of F_2 for fixed values of x, and the comparison of this variation to a model evolved with the Altarelli-Parisi equations. The upper six data sets have been multiplied by the indicated factors to separate them on the plot. The data were compiled by M. Virchaux and R. Voss for the Particle Data Group, *Phys. Rev.* **D50**, 1173 (1994), Fig. 32.2. The complete references to the original experiments are given there.

which share its total momentum, as its wavefunction is probed on finer and finer distance scales.

Figure 17.22 shows the evolution of the combination of distribution functions that is measured in deep inelastic scattering, as a function of Q^2. We see the characteristic decrease of the distribution functions at large x and the increase at small x. The data are compared to a model evolved according to the Altarelli-Parisi equations; this model apparently describes the data quite well.

17.6 Measurements of α_s

Before concluding our introductory survey of QCD, we should summarize the quantitative verification of the theory. We discussed precision tests of QED in Section 6.3, bringing together various measurements of the coupling α; the best determinations agree to eight significant figures. Since QCD perturbation theory works only for hard-scattering processes, with uncertainties due to soft processes that are difficult to estimate, this theory has not been tested to such extreme accuracy. Nevertheless, it is interesting to bring together the best available determinations of α_s, to see how well they agree.

In order to compare values of α_s, it is necessary to express these using a common set of conventions. First, one must set the renormalization scale; a useful choice is the mass of the neutral weak boson Z^0: $m_Z = 91.19$ GeV. Second, one must fix the renormalization scheme that defines the QCD coupling constant at this scale. It has become conventional to use as a standard the bare coupling after regularization by modified minimal subtraction, Eq. (11.77). The resulting standard coupling constant is called $\alpha_{s\overline{\rm MS}}(m_Z^2)$.

Measurements of α_s from a number of types of experiments are summarized in Table 17.1. In Section 17.2 we saw that one can obtain a value of α_s from the measurement of the total cross section for e^+e^- annihilation to hadrons or, equivalently, the ratio R of the number of observed hadronic and leptonic events. An independent measurement of α_s can be obtained from the fraction of e^+e^- annihilation events with three-jet final states or, equivalently, from the transverse momentum distribution of produced hadrons relative to the jet axis. A number of measurements of this type are collected and averaged under the heading 'Event shapes'. A similar measurement of α_s is obtained from the measurement of the transverse momentum spectrum of W bosons produced from quark-antiquark annihilation at high-energy $p\bar{p}$ colliders. The gluon radiative correction to the vertex in deep inelastic neutrino scattering can also be used to extract α_s. The rate of Bjorken scaling violation in deep inelastic scattering is controlled by α_s, and so this effect provides another α_s measurement. The decays of the lightest $b\bar{b}$ bound state Υ and the $c\bar{c}$ bound state ψ are governed by QCD and yield a measurement of α_s. Finally, the spectrum of $c\bar{c}$ and $b\bar{b}$ bound states can be computed numerically in terms of the QCD coupling constant, and the comparison with experiment gives a determination of α_s.

Table 17.1. Values of $\alpha_s(m_Z)$ Obtained from QCD Experiments

Process:	$\alpha_s(m_Z)$	Q (GeV)
Deep inelastic scattering	0.118 (6)	1.7
R in τ lepton decay	0.123 (4)	1.8
ψ, Υ spectroscopy	0.110 (6)	2.3
Transverse momentum of W production	0.121 (24)	4.
Deep inelastic scattering (evolution)	0.112 (4)	5.
Event shapes in e^+e^- annihilation	0.121 (6)	5.8,9.1
Rate for ψ, Υ decay	0.108 (10)	9.5
R in e^+e^- annihilation (20–65 GeV)	0.124 (21)	35.
R in Z^0 decay	0.124 (7)	91.2

The values of $\alpha_s(m_Z)$ displayed in this table are obtained by fitting experimental results to the theoretical expressions given by perturbative QCD using minimal subtraction. The values of α_s have been evolved to $Q = m_Z$ using the renormalization group equation. R refers to the ratio of cross sections or partial widths to hadrons versus leptons. The numbers in parentheses are the standard errors in the last displayed digits. The column labelled 'Q' gives an idea of the value of Q at which the measurement was made. (Typically, these measurements average over a range of Q, and that averaging is taken into account in the quoted values of α_s.) This table is based on the results compiled by I. Hinchliffe in his article for the Particle Data Group, *Phys. Rev.* **D50**, 1297 (1994). This article contains a full set of references and a discussion of the sources of uncertainty in these determinations.

The table shows the values of α_s extracted from each of these measurements, expressed in terms of the value in the reference conventions, $\alpha_{s\overline{\text{MS}}}(m_Z^2)$. We see that several of the experiments determine α_s to an accuracy of 5%, and that the various determinations are consistent with one another at this level. In Fig. 17.23, we have plotted the original values of α_s represented in Table 17.1, before conversion to a common scale, versus the momentum scale Q at which each was obtained. This comparison gives a striking direct verification of the running of α_s.

At the beginning of this chapter, we wrote down a candidate for the fundamental theory of strong interactions using only a few simple principles: the existence of quarks and the identification of their quantum numbers, and the idea that the theory of the quark interactions should be an asymptotically free gauge theory. It is remarkable that these simple considerations have led us to a description of strong interactions that is quantitatively correct for a broad range of phenomena in the hard-scattering regime where asymptotic freedom can be used as a tool for calculation.

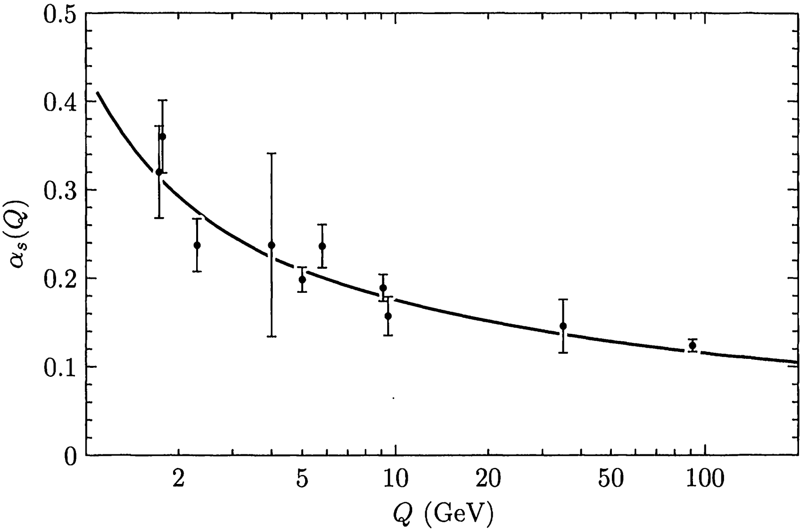


Figure 17.23. Measurements of α_s, plotted against the momentum scale Q at which the measurement was made. This figure was constructed by evolving the values of $\alpha_s(m_Z)$ listed in Table 17.1 back to the values of Q indicated in the table. The value for e^+e^- event shapes has been split into two points corresponding to experiments at the TRISTAN and LEP accelerators. These values are compared to the theoretical expectation from the renormalization group evolution with the initial condition $\alpha_s(m_Z) = 0.117$.

Problems

17.1 Two-loop renormalization group relations.

(a) In higher orders of perturbation theory, the expression for the QCD β function will be a series

$$\beta(g) = -\frac{b_0}{(4\pi)^2}g^3 - \frac{b_1}{(4\pi)^4}g^5 - \frac{b_2}{(4\pi)^6}g^7 + \cdots.$$

Integrate the renormalization group equation and show that the running coupling constant is now given by

$$\alpha_s(Q^2) = \frac{4\pi}{b_0}\left[\frac{1}{\log(Q^2/\Lambda^2)} - \frac{b_1}{b_0^2}\frac{\log\log(Q^2/\Lambda^2)}{(\log(Q^2/\Lambda^2))^2} + \cdots\right],$$

where the omitted terms decrease as $(\log(Q^2/\Lambda^2))^{-2}$.

(b) Combine this formula with the perturbation series for the e^+e^- annihilation cross section:

$$\sigma(e^+e^- \to \text{hadrons}) = \sigma_0 \cdot \left(3\sum_f Q_f^2\right) \cdot \left[1 + \frac{\alpha_s}{\pi} + a_2\left(\frac{\alpha_s}{\pi}\right)^2 + \mathcal{O}(\alpha_s^3)\right].$$

The coefficient a_2 depends on the details of the renormalization conditions defining α_s. Show that the leading two terms in the asymptotic behavior of $\sigma(s)$ for large s depend only on b_0 and b_1 and are independent of a_2 and b_2. Thus the first *two* coefficients of the QCD β function are independent of the renormalization prescription.

17.2 A direct test of the spin of the gluon. In this problem, we compare the predictions of QCD with those of a model in which the interaction of quarks is mediated by a scalar boson. Let the coupling of the scalar gluon to quarks be given by

$$\delta\mathcal{L} = gS\bar{q}q,$$

and define $\alpha_g = g^2/4\pi$.

(a) Using the technique described in parts (b) and (c) of the Final Project of Part I, compute the cross section for $e^+e^- \to q\bar{q}S$ to the leading order of perturbation theory. This cross section depends on the energies of the q, $\bar{q}$, and S, which we represent as fractions x_1, x_2, x_3 of the electron beam energy, as in Eq. (17.18). Show that

$$\frac{d^2\sigma}{dx_1 dx_2}(e^+e^- \to q\bar{q}S) = \frac{4\pi\alpha^2 Q_q^2}{3s} \cdot \frac{\alpha_g}{4\pi} \frac{x_3^2}{(1-x_q)(1-x_{\bar{q}})}.$$

(b) In practice, it is very difficult to tell quarks from gluons experimentally, since both particles appear as jets of hadrons. Therefore, let x_a be the largest of x_1, x_2, x_3, let x_b be the second largest, and let x_c be the smallest. Sum over the various possibilities to derive an expression for $d^2\sigma/dx_a dx_b$, both in QCD, using Eq. (17.18), and in the scalar gluon model. Show that these models can be distinguished by their distributions in the x_a, x_b plane.

17.3 Quark-gluon and gluon-gluon scattering.

(a) Compute the differential cross section

$$\frac{d\sigma}{d\hat{t}}(q\bar{q} \to gg)$$

for quark-antiquark annihilation in QCD to the leading order in α_s. This is most easily done by computing the amplitudes between states of definite quark and gluon helicity. Ignore all masses. Use explicit polarization vectors and spinors, for example,

$$\epsilon^\mu = \frac{1}{\sqrt{2}}(0, 1, i, 0)$$

for a right-handed gluon moving in the $+\hat{3}$ direction. You need only consider transversely polarized gluons. By helicity conservation, only the initial states $q_L\bar{q}_R$ and $q_R\bar{q}_L$ can contribute; by parity, these two states give identical cross sections. Thus it is necessary only to compute the amplitudes for the three processes

$$q_L\bar{q}_R \to g_R g_R,$$
$$q_L\bar{q}_R \to g_R g_L,$$
$$q_L\bar{q}_R \to g_L g_L.$$

In fact, by CP invariance, the first and third processes have equal cross sections. After computing the amplitudes, square them and combine them properly with

color factors to construct the various helicity cross sections. Finally, combine these to form the total cross section averaged over initial spins and colors.

(b) Compute the differential cross section

$$\frac{d\sigma}{d\hat{t}}(gg \to gg)$$

for gluon-gluon scattering. There are 16 possible combinations of helicities, but many of them are related to each other by parity and crossing symmetry. All 16 can be built up from the three amplitudes for

$$g_R g_R \to g_R g_R,$$
$$g_R g_R \to g_R g_L,$$
$$g_R g_R \to g_L g_L.$$

Show that the last two of these amplitudes vanish. The first can be dramatically simplified using the Jacobi identity. When the smoke clears, only three of the 16 polarized gluon scattering cross sections are nonzero. Combine these to compute the spin- and color-averaged differential cross section.

17.4 The gluon splitting function. Compute the gluon splitting function (17.130) for the Altarelli-Parisi equations. To carry out this computation, first compute the matrix elements of the three-gluon vertex shown in Fig. 17.20 between gluon states of definite helicity. Combine these to derive the splitting function in the region $x < 1$. Then fix the singularity of the splitting function at $x = 1$ to give this function the correct overall normalization.

17.5 Photoproduction of heavy quarks. Consider the process of heavy quark pair photoproduction, $\gamma + p \to Q\bar{Q} + X$, for a heavy quark of mass M and electric charge Q. If M is large enough, any diagram contributing to this process must involve a large momentum transfer; thus a perturbative QCD analysis should apply. This idea applies in practice already for the production of c quark pairs. Work out the cross section to the leading order in QCD. Choose the parton subprocess that gives the leading contribution to this reaction, and write the parton-model expression for the cross section. You will need to compute the relevant subprocess cross section, but this can be taken directly from one of the QED calculations in Chapter 5. Then use this result to write an expression for the cross section for γ-proton scattering.

17.6 Behavior of parton distribution functions at small x. It is possible to solve the Altarelli-Parisi equations analytically for very small x, using some physically motivated approximations. This discussion is based on a paper of Ralston.†

(a) Show that the Q^2 dependence of the right-hand side of the A-P equations can be expressed by rewriting the equations as differential equations in

$$\xi = \log\log\left(\frac{Q^2}{\Lambda^2}\right),$$

where Λ is the value of Q^2 at which $\alpha_s(Q^2)$, evolved with the leading-order β function, formally goes to infinity.

†J. P. Ralston, *Phys. Lett.* **172B**, 430 (1986).

(b) Since the branching functions to gluons are singular as z^{-1} as $z \to 0$, it is reasonable to guess that the gluon distribution function will blow up approximately as x^{-1} as $x \to 0$. The resulting distribution

$$dx\, f_g(x) \sim \frac{dx}{x}$$

is approximately scale invariant, and so its form should be roughly preserved by the A-P equations. Let us, then, make the following two approximations: (1) the terms involving the gluon distribution completely dominate the right-hand sides of the A-P equations; and (2) the function

$$\tilde{g}(x, Q^2) = x f_g(x, Q^2)$$

is a slowly varying function of x. Using these approximations, and the limit $x \to 0$, show that the A-P equation for $f_g(x)$ can be converted to the following differential equation:

$$\frac{\partial^2}{\partial w \partial \xi}\tilde{g}(x, \xi) = \frac{12}{b_0}\, \tilde{g}(x, \xi),$$

where $w = \log(1/x)$ and $b = (11 - \frac{2}{3}n_f)$. Show that if $w\xi \gg 1$, this equation has the approximate solution

$$\tilde{g} = K(Q^2) \cdot \exp\left(\left[\frac{48}{b_0} w(\xi - \xi_0)\right]^{1/2}\right),$$

where $K(Q^2)$ is an initial condition.

(c) The quark distribution at very small x is mainly created by branching of gluons. Using the approximations of part (b), show that, for any flavor of quark, the right-hand side of the A-P equation for $f_q(x)$ can be approximately integrated to yield an equation for $\tilde{q}(x) = x f_q(x)$:

$$\frac{\partial}{\partial \xi}\tilde{q}(x, \xi) = \frac{2}{3b_0}\tilde{g}(x, \xi).$$

Show, again using $w\xi \gg 1$, that this equation has as its integral

$$\tilde{q} = \left(\frac{\xi - \xi_0}{27 b_0 w}\right)^{1/2} K(Q^2) \cdot \exp\left(\left[\frac{48}{b_0} w(\xi - \xi_0)\right]^{1/2}\right).$$

(d) Ralston suggested that the initial condition

$$K(Q^2) = 50.36(\exp(\xi - \xi_0) - 0.957) \cdot \exp\left[-7.597(\xi - \xi_0)^{1/2}\right],$$

with $Q_0^2 = 5\ \text{GeV}^2$, $\Lambda = 0.2$ GeV, and $n_f = 5$, gave a reasonable fit to the known properties of parton distributions, extrapolated into the small x region. Use this function and the results above to sketch the behavior of the quark and gluon distributions at small x and large Q^2.

Chapter 18

Operator Products and Effective Vertices

Our analysis of QCD in Chapter 17 was founded on the principle of asymptotic freedom, which told us that strong interaction processes with large momentum transfer might reliably be treated in weak-coupling perturbation theory. So far, however, we have made little use in QCD of the more powerful tools of the renormalization group. In this chapter, we will work out some implications of the Callan-Symanzik equation in QCD. We will see that asymptotically free theories have their own characteristic scaling behavior, with corrections in the form of anomalous powers of logarithms of the momentum scale. Though these corrections are generally weaker than those in the scalar field theories studied in Chapter 13, they nevertheless have important qualitative effects on the strong interactions.

We begin by considering the scaling law for mass terms in QCD, taking over directly the formalism that we used to describe the mass term of ϕ^4 theory in Sections 12.4 and 12.5. Other applications, however, require a more powerful theoretical tool, the *operator product expansion.* Section 18.3 introduces a general description of products of operators in quantum field theory and explains how such operator products are constrained by the Callan-Symanzik equation. The last two sections use this tool to develop a new viewpoint toward deep inelastic scattering and other hard processes in QCD.

18.1 Renormalization of the Quark Mass Parameter

Up to this point, we have always assumed that quark masses are small enough that they can be ignored in high-energy processes. This is not always an adequate assumption even for the light quarks u, d, s; for the heavier quarks c, b, t, the masses can have very important effects. However, since isolated quarks do not exist, it is not possible to define the mass of a quark unambiguously. In the discussion to follow, we will consider the quark mass to be a parameter of QCD perturbation theory, defined by a renormalization prescription at some renormalization scale M.

Because we define the quark mass as we would a coupling constant, by a renormalization convention, we should expect that this parameter will run according to a renormalization group evolution, so that different values of the

mass parameter apply to different processes. We say that our original prescription leads to an *effective* quark mass, which depends on the momentum scale at which it is evaluated. In this section, we will work out the leading dependence of this effective mass on the momentum scale.

The basic formalism for effective mass terms was set out in Section 12.5. To add a mass term to the QCD Lagrangian, we must first define the mass operator $(\bar{q}q)$ by a renormalization prescription at a scale M. Then we can define the quark mass by adding to the Lagrangian the term

$$\Delta\mathcal{L}_m = -m(\bar{q}q)_M. \tag{18.1}$$

In this discussion, we will assume that the quark mass m is small enough that we need only keep terms of leading order in m. We will also assume, for simplicity, that we have such a mass term for only one quark flavor.

In the zero-mass limit, Green's functions of the operator $(\bar{q}q)$ with quark fields,

$$\begin{aligned} G^{(n,k)}&(x_1,\ldots,x_n,y_1,\ldots,y_n,z_1,\ldots,z_k) \\ &= \langle q(x_1)\cdots q(x_n)\bar{q}(y_1)\cdots\bar{q}(y_n)\bar{q}q(z_1)\cdots\bar{q}q(z_k)\rangle\,, \end{aligned} \tag{18.2}$$

obey the Callan-Symanzik equation

$$\left[M\frac{\partial}{\partial M}+\beta\frac{\partial}{\partial g}+2n\gamma+k\gamma_{\bar{q}q}\right]G^{(n,k)}(\{x_i\},\{y_i\},\{z_j\},g,M)=0, \tag{18.3}$$

where γ is the anomalous dimension of the quark field and $\gamma_{\bar{q}q}$ is the anomalous dimension of the operator $\bar{q}q$. If we include the mass terms in the Lagrangian according to (18.1), the Green's function of n quark fields and n antiquark fields satisfies

$$\left[M\frac{\partial}{\partial M}+\beta\frac{\partial}{\partial g}+2n\gamma+\gamma_{\bar{q}q}m\frac{\partial}{\partial m}\right]G^{(n)}(\{x_i\},\{y_i\},g,m,M)=0. \tag{18.4}$$

The derivative with respect to m counts the number of times the mass operator is used. In Section 12.5, we traded the variable m, with the dimensions of mass, for a dimensionless variable. However, in QCD, it is just as convenient to consider the dimensionful parameter m as a coupling constant. The solution of the Callan-Symanzik equation will then contain a running mass parameter $\overline{m}(Q)$, which depends on a typical momentum Q of the Green's function. This parameter is defined as the solution to a renormalization group equation analogous to Eq. (12.126). For this case, the equation is

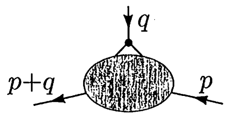
$$\frac{d}{d\log(Q/M)}\overline{m} = \gamma_{\bar{q}q}(\bar{g})\cdot\overline{m}, \tag{18.5}$$

with the initial condition

$$\overline{m}(M) = m. \tag{18.6}$$

The quantity $\overline{m}(Q)$ is the *effective mass*, which should be used to compute the mass effects on quark production or scattering processes with the momentum transfer Q.

To compute $\overline{m}(Q)$ explicitly, we need to work out the anomalous dimension of the mass operator $\gamma_{\bar{q}q}$. This can be done as explained in Section 12.4. We define the normalization of the operator explicitly by the prescription that the vertex function of $(\bar{q}q)$ between renormalized quark fields should satisfy



$$= 1 \qquad (18.7)$$

for $p^2 = q^2 = (p+q)^2 = -M^2$. To preserve (18.7), we will need a counterterm vertex $\delta_{\bar{q}q}$ with the structure of the operator insertion. Then, as in Eq. (12.112), the anomalous dimension is given to one-loop order by

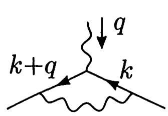
$$\gamma_{\bar{q}q} = M\frac{\partial}{\partial M}\big(-\delta_{\bar{q}q} + \delta_2\big), \qquad (18.8)$$

where δ_2 is the counterterm for the quark field strength renormalization, defined in Fig. 16.8. Correlation functions of the gauge-invariant operator $(\bar{q}q)$ are gauge invariant, and so the various terms in the Callan-Symanzik equation for this function must sum to a gauge-invariant result. Since the leading coefficient of $\beta(g)$ is independent of the gauge and of other conventions, it follows from (18.3) that the leading coefficient of $\gamma_{\bar{q}q}$ is also convention independent. The counterterms δ_2 and $\delta_{\bar{q}q}$ both depend on the gauge. This argument shows that the gauge dependence must cancel in (18.8). In the calculation to follow, and in the other anomalous dimension calculations in this chapter, we will work consistently in Feynman-'t Hooft gauge.

We have already computed the divergent part of the counterterm δ_2 in Feynman-'t Hooft gauge in Section 16.4. Evaluating the group-theory factor in the result (16.77) for QCD, we find

$$\delta_2 = -\frac{4}{3}\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \qquad (18.9)$$

To compute $\delta_{\bar{q}q}$, we must work out the one-loop correction to the vertex (18.7). This is given by the diagram



$$= \int \frac{d^4k}{(2\pi)^4}\,(ig)^2 t^a\gamma^\mu \frac{i(\not{k}+\not{q})}{(k+q)^2}\cdot 1\cdot \frac{i\not{k}}{k^2}t^a\gamma_\mu \frac{-i}{(k-p)^2}. \qquad (18.10)$$

In the expression for this diagram, the factor 1 represents the $\bar{q}q$ operator insertion. In the corresponding diagram for the renormalization of the quark number current $j^\nu = \bar{q}\gamma^\nu q$, this factor would be replaced by γ^ν. Since we need only the divergent part of the vertex renormalization (18.10), we can

approximate the integrand by its value for large k. Then this diagram becomes



$$\begin{aligned}
&\sim \int \frac{d^4k}{(2\pi)^4} (ig)^2 t^a \gamma^\mu \frac{i\not{k}}{k^2} \cdot 1 \cdot \frac{i\not{k}}{k^2} t^a \gamma_\mu \frac{-i}{k^2} \\
&\sim -i\frac{4}{3}g^2 \int \frac{d^4k}{(2\pi)^4} \frac{d \cdot k^2}{(k^2)^3} \qquad (18.11)\\
&\sim \frac{4}{3}g^2 \cdot 4 \cdot \frac{1}{(4\pi)^2}\Gamma(2-\tfrac{d}{2}).
\end{aligned}$$

To preserve the normalization condition (18.7), we must add the counterterm

$$\delta_{\bar{q}q} = -4 \cdot \frac{4}{3}\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \qquad (18.12)$$

Assembling (18.8), (18.9), and (18.12), we find

$$\gamma_{\bar{q}q} = -8\frac{g^2}{(4\pi)^2}. \qquad (18.13)$$

As we have noted in the previous paragraph, the anomalous dimension γ_j of the quark number current can be found by a very similar calculation. This will give a good check on our formalism, since, as we have argued above Eq. (12.110), a conserved current is unambiguously normalized by its integral, the conserved charge, and so must have zero anomalous dimension. If we substitute γ^ν for 1 in (18.10) and use the same set of approximations to reduce the integral, we find in the numerator the Dirac matrix structure

$$\begin{aligned}
\gamma^\mu \not{k} \gamma^\nu \not{k} \gamma_\mu &= \frac{1}{d}k^2 \gamma^\mu \gamma^\lambda \gamma^\nu \gamma_\lambda \gamma_\mu \\
&= \frac{1}{4}(-2)^2 k^2 \gamma^\nu.
\end{aligned} \qquad (18.14)$$

Then, instead of (18.12), we need the counterterm

$$\delta_j = -\frac{4}{3}\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \qquad (18.15)$$

Combining this result with (18.9), we find

$$\gamma_j = 0, \qquad (18.16)$$

in accord with our general arguments.

If we replace the gamma function in (18.11) by an explicit factor of $\log(\Lambda^2/Q^2)$, and then subtract the divergence using the counterterm (18.12), we find that the vertex diagram behaves as



$$= \frac{4}{3} \cdot 4\frac{g^2}{(4\pi)^2} \log \frac{M^2}{Q^2}. \qquad (18.17)$$

This diagram gives an enhancement at small external momenta. Some of this enhancement is associated with the (gauge-dependent) rescaling of the external quark fields; relation (18.8) tells us how to extract the piece of this

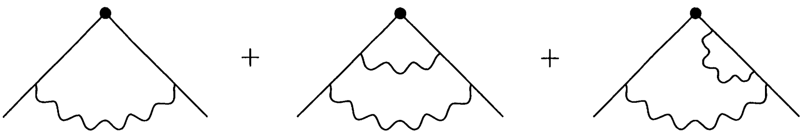


Figure 18.1. Diagrams giving the leading logarithmic contributions to the momentum dependence of the quark effective mass.

logarithm associated with the gauge-invariant enhancement of the effective mass. Thus, to order α_s,

$$\overline{m}(Q) = m \cdot \left(1 + 4\frac{g^2}{(4\pi)^2} \log \frac{M^2}{Q^2}\right). \tag{18.18}$$

To compute the momentum dependence of the effective mass more accurately, we must take two more features of the calculation into account. First, the quantity $(\alpha_s \log(M^2/Q^2))$ may become of order 1, and, in this case, we must take into account all leading logarithmic terms of the form $(\alpha_s \log(M^2/Q^2))^n$. Contributions of this type come from all the diagrams shown in Fig. 18.1. Second, the coupling constant α_s is itself a function of the momentum scale, giving a further enhancement to contributions from small Q. Both of these effects are properly accounted by solving the renormalization group equation (18.5). To the leading order in g^2, this equation takes the explicit form

$$\frac{d}{d\log(Q/M)}\overline{m} = -8\frac{\bar{g}^2}{(4\pi)^2}m = -2\frac{\alpha_s(Q^2)}{\pi}\overline{m}. \tag{18.19}$$

Inserting the solution of the renormalization group equation for $\bar{g}$ in the form (17.17), we find

$$\frac{d}{d\log(Q/M)}\overline{m} = -\frac{8}{b_0 \log(Q^2/\Lambda^2)}\overline{m}, \tag{18.20}$$

where b_0 is the first coefficient of the QCD β function and Λ is now the QCD scale parameter defined in (17.16). The integral of this equation, satisfying the initial condition (18.6), is

$$\overline{m}(Q^2) = \left(\frac{\log(M^2/\Lambda^2)}{\log(Q^2/\Lambda^2)}\right)^{4/b_0} m. \tag{18.21}$$

Recall that $b_0 = 11 - \frac{2}{3}n_f$ in QCD. Another way to express (18.21) is by writing

$$\overline{m}(Q^2) = \left(\frac{\alpha_s(Q^2)}{\alpha_s(M^2)}\right)^{4/b_0} m. \tag{18.22}$$

Just as an illustration, take $n_f = 4$ and $\Lambda = 150$ MeV; then the effective masses of the light quarks increase by about a factor 2 from $Q = 100$ GeV to $Q = 1$ GeV.

The method we have just used for computing the QCD enhancement of the quark mass operator applies equally well to the matrix elements of any other gauge-invariant operator. We conclude this section by recapitulating the conclusions of the argument in their more general form.

Let $\mathcal{O}(x)$ be any gauge-invariant operator in QCD. As we saw for the mass term, the one-loop corrections to the matrix elements of this operator may contain enhancement or suppression terms proportional to $\alpha_s \log(M^2/Q^2)$, where Q is the momentum scale of a QCD process mediated by $\mathcal{O}(x)$ and M is the renormalization scale used to define the operator normalization. The part of these one-loop corrections specifically associated with the operator normalization is given by the anomalous dimension $\gamma_{\mathcal{O}}$. For an operator containing n quark or antiquark fields and k gluon fields,

$$\gamma_{\mathcal{O}} = M\frac{\partial}{\partial M}\Big(-\delta_{\mathcal{O}} + \frac{n}{2}\delta_2 + \frac{k}{2}\delta_3\Big), \tag{18.23}$$

where $\delta_{\mathcal{O}}$ is the counterterm needed to preserve the operator normalization condition and δ_2 and δ_3 are the counterterms for the quark and gluon field strength renormalization defined in Fig. 16.8. From (18.23), we can derive the explicit one-loop expression for $\gamma_{\mathcal{O}}$ in the form

$$\gamma_{\mathcal{O}} = -a_{\mathcal{O}}\frac{g^2}{(4\pi)^2}. \tag{18.24}$$

Using this result, we can solve the renormalization group equation for the coefficient of $\mathcal{O}(x)$ and find the QCD renormalization factor

$$\left(\frac{\log(M^2/\Lambda^2)}{\log(Q^2/\Lambda^2)}\right)^{a_{\mathcal{O}}/2b_0}, \tag{18.25}$$

where b_0 is the first coefficient of the QCD β function,

$$b_0 = 11 - \frac{2}{3}n_f. \tag{18.26}$$

The QCD renormalization (18.25) is an enhancement at small momenta if $a_{\mathcal{O}} > 0$.

In the remainder of this chapter, we will present further examples of this enhancement or suppression by QCD logarithms. In many cases, we will see that these factors lead to striking and nontrivial physical effects.

18.2 QCD Renormalization of the Weak Interaction

Our next example of the appearance of QCD enhancement factors occurs in the theory of the weak interactions of hadrons. In Section 17.3, we introduced the weak interaction coupling of quarks and leptons, which we described by an effective Lagrangian. For our analysis here, we will need to know a few more details of the structure of the weak interactions, so we begin this section by presenting these facts. The complete structure of the weak interactions of quarks and leptons will be discussed systematically in Chapter 20.

As we discussed in Section 17.3, the weak interactions among quarks and leptons are described by an effective Lagrangian resulting from the exchange of a virtual W vector boson. In (17.31), we wrote the effective vertex that couples quarks to leptons:

$$\Delta\mathcal{L} = \frac{g^2}{2m_W^2}\bar{\ell}\gamma^\mu\frac{(1-\gamma^5)}{2}\nu\bar{u}\gamma_\mu\frac{(1-\gamma^5)}{2}d + \text{h.c.} \tag{18.27}$$

In this chapter, we will mainly be concerned with the effects of this interaction for momentum scales much larger than 1 GeV. Thus, we will ignore quark masses. All fermion fields that appear in the weak-interaction vertices are multiplied by the left-handed projector $\frac{1}{2}(1-\gamma^5)$. In the rest of this section, we will not write this projector explicitly; rather, we will denote the projection by a subscript L. We will also introduce the Fermi constant, given by (17.32). Then (18.27) can be rewritten as

$$\Delta\mathcal{L} = \frac{4G_F}{\sqrt{2}}\big(\bar{\ell}_L\gamma^\mu\nu_L\big)\big(\bar{u}_L\gamma_\mu d_L\big) + \text{h.c.} \tag{18.28}$$

There is an analogous vertex that represents W exchange between pairs of quarks; this has the form

$$\Delta\mathcal{L} = \frac{4G_F}{\sqrt{2}}\big(\bar{d}_L\gamma^\mu u_L\big)\big(\bar{u}_L\gamma_\mu d_L\big) + \text{h.c.} \tag{18.29}$$

However, for the discussion of this chapter, we will need to write a modified, and less approximate, expression. When we discuss the theory of weak interactions in detail in Chapter 20, we will learn that the charge $+2/3$ quarks (u, c, t) couple to the charge $-1/3$ quarks (d, s, b) through the weak interactions via a unitary rotation. Thus, for example, u couples to the combination

$$\cos\theta_c\, d + \sin\theta_c\, s, \tag{18.30}$$

plus a small admixture of b, which we will ignore in this section. The mixing angle θ_c is called the *Cabibbo angle*. Because of this rotation, the weak interaction effective Lagrangian coupling quarks to quarks actually contains a number of terms, of which a particularly important one is

$$\Delta\mathcal{L} = \frac{4G_F}{\sqrt{2}}\cos\theta_c\sin\theta_c(\bar{d}_L\gamma^\mu u_L)(\bar{u}_L\gamma_\mu s_L). \tag{18.31}$$

This term allows the s quarks to decay through the process $s \to u\bar{u}d$. Similarly, the rotation of (18.28) produces the effective interaction

$$\Delta\mathcal{L} = \frac{4G_F}{\sqrt{2}} \sin\theta_c (\bar{\ell}_L \gamma^\mu \nu_L)(\bar{u}_L \gamma_\mu s_L), \tag{18.32}$$

which leads to the decay $s \to u\bar{\nu}\ell$. These weak interaction processes are referred to as *nonleptonic* and *semileptonic* decay processes, respectively. Similar expressions apply to the other heavy quarks.

Given that (18.31) and (18.32) describe the weak interaction coupling of the s quark at a fundamental level, we now discuss the modification of these couplings by QCD logarithms. We have seen in the previous section that QCD corrections have a profound effect in enhancing the strength of the quark mass term of the underlying Lagrangian. We will now investigate whether the strength of the weak interactions can receive a similar enhancement.

We first consider the semileptonic weak interaction operator (18.32). The leptonic fermion bilinear is not affected by QCD, so the QCD enhancement of this operator is just the same as that of its quark component

$$\bar{u}_L \gamma_\mu s_L. \tag{18.33}$$

However, this operator is a current and so has $\gamma = 0$. In terms of diagrams, the logarithmic enhancement resulting from the diagram shown in Fig. 18.2 is canceled by the quark field-strength renormalization, as we saw already in our discussion of the current vertex is Section 18.1. The left-handed projector $\frac{1}{2}(1-\gamma^5)$ commutes through the diagram and has no effect on the final result. The same remark applies to the semileptonic weak interaction that links u and d quarks. It implies, for that case, that the normalization of the cross sections for deep inelastic neutrino scattering given in (17.35) is not affected by QCD logarithms.

In the case of nonleptonic weak interactions, however, the effect of QCD is not so simple. Let us first compute the Feynman diagrams that give the leading corrections to the renormalization of the weak interaction vertex (18.31) and then, at a later stage, build up the renormalization group interpretation of these results.

At order α_s, the nonleptonic weak interaction vertex receives corrections from the diagrams shown in Fig. 18.3. Notice that the first diagram is precisely the current renormalization found in the semileptonic case. The second diagram gives the analogous renormalization of the second quark current. In the computation of γ, these two contributions cancel the contributions from the field-strength renormalization of the four quark fields. The remaining four diagrams of Fig. 18.3 are new contributions which contribute potentially large rescaling factors.

We now compute these diagrams, beginning with the third diagram in Fig. 18.3. As in the computation of Section 18.1, we are interested in the logarithmically divergent contribution associated with values of the loop momentum k much larger than the external momenta. The simplest way to extract this

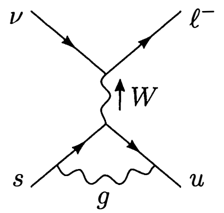


Figure 18.2. QCD correction to the strength of the semileptonic weak interaction vertex.

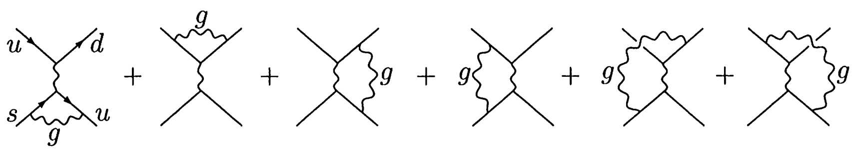


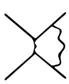
Figure 18.3. QCD corrections to the strength of the nonleptonic weak interaction vertex.

contribution is to compute each diagram in the approximation of zero external momentum. In writing the expression for these diagrams, we will omit the prefactor

$$\frac{4G_F}{\sqrt{2}}\cos\theta_c \sin\theta_c. \tag{18.34}$$

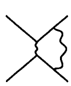
We will retain the quark fields to represent the external states, so that our final expressions will have the form of rescaled operators.

Using this notation, the third diagram in Fig. 18.3 has the value



$$= \int \frac{d^4k}{(2\pi)^4}(ig)^2\frac{-i}{k^2}\Big(\bar{d}_L\gamma^\nu t^a \frac{i\not{k}}{k^2}\gamma^\mu u_L\Big)\Big(\bar{u}_L\gamma_\nu t^a \frac{-i\not{k}}{k^2}\gamma_\mu s_L\Big). \tag{18.35}$$

Using the symmetry of the k integral, we extract the divergent piece:



$$\begin{aligned} &= ig^2\int \frac{d^4k}{(2\pi)^4}\frac{k^2/d}{(k^2)^3}\big(\bar{d}_L\gamma^\nu t^a\gamma^\lambda\gamma^\mu u_L\big)\big(\bar{u}_L\gamma_\nu t^a\gamma_\lambda\gamma_\mu s_L\big) \\ &= -\frac{g^2}{4}\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^2}\big(\bar{d}_L\gamma^\nu t^a\gamma^\lambda\gamma^\mu u_L\big)\big(\bar{u}_L\gamma_\nu t^a\gamma_\lambda\gamma_\mu s_L\big). \end{aligned} \tag{18.36}$$

To put the product of quark fields into a more familiar form, we apply the Fierz transformation discussed at the end of Section 3.4. If the color matrices t^a were not present, the product of fermion fields would be exactly the one appearing in (3.82), and we would find

$$\big(\bar{d}_L\gamma^\nu\gamma^\lambda\gamma^\mu u_L\big)\big(\bar{u}_L\gamma_\nu\gamma_\lambda\gamma_\mu s_L\big) = 16\bar{d}_L\gamma^\nu u_L\bar{u}_L\gamma_\nu s_L. \tag{18.37}$$

The matrices t^a redirect the color quantum numbers of the quark fields. To clarify this, we need the analogue of identity (3.77) for color. To find this

identity, consider the color invariant

$$(t^a)_{ij}(t^a)_{k\ell}. \tag{18.38}$$

The indices i, k transform according to the 3 representation of color; the indices j, ℓ transform according to the $\bar{3}$. Thus, (18.38) must be a linear combination of the two possible ways to contract these indices,

$$A\delta_{i\ell}\delta_{kj} + B\delta_{ij}\delta_{k\ell}. \tag{18.39}$$

The constants A and B can be determined by contracting (18.38) and (18.39) with δ_{ij} and with δ_{jk} and adjusting A and B so that the contractions of (18.39) obey the identities

$$\mathrm{tr}[t^a](t^a)_{k\ell} = 0; \qquad (t^a t^a)_{i\ell} = \tfrac{4}{3}\delta_{i\ell}. \tag{18.40}$$

This gives the identity

$$(t^a)_{ij}(t^a)_{k\ell} = \tfrac{1}{2}\big(\delta_{i\ell}\delta_{kj} - \tfrac{1}{3}\delta_{ij}\delta_{k\ell}\big). \tag{18.41}$$

A similar relation holds for the generators of $SU(N)$ in the fundamental representation, with $(1/3)$ replaced by $(1/N)$ in that case.

Inserting (18.41) into (18.36), we find that the first term of the identity generates a new four-fermion operator,

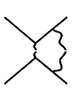
$$\big(\bar{d}_{Li}\gamma^\nu\gamma^\lambda\gamma^\mu u_{Lj}\big)\big(\bar{u}_{Lj}\gamma_\nu\gamma_\lambda\gamma_\mu s_{Li}\big), \tag{18.42}$$

where i, j are color indices. Applying the Fierz rearrangement in (18.37), and then applying the additional rearrangement (3.79), we can convert this operator to the form

$$16(\bar{d}_{Li}\gamma^\nu u_{Lj})(\bar{u}_{Lj}\gamma_\nu s_{Li}) = 16(\bar{u}_{Lj}\gamma^\nu u_{Lj})(\bar{d}_{Li}\gamma_\nu s_{Li}). \tag{18.43}$$

The minus sign in (3.79) is compensated by a minus sign from interchanging the order of fermion fields. The final result is a product of color-singlet quark currents; however, the fields in these currents are associated differently from the original operator.

The final result of our evaluation of this diagram is



$$= -4g^2\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^2}\big[\tfrac{1}{2}\bar{u}_L\gamma^\mu u_L\bar{d}_L\gamma_\mu s_L - \tfrac{1}{6}\bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L\big]. \tag{18.44}$$

The fourth diagram of Fig. 18.3 gives precisely the same contribution.

The evaluation of the last two diagrams in Fig. 18.3 is quite similar. The fifth diagram gives

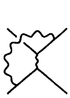


$$\begin{aligned} &= \int\frac{d^4k}{(2\pi)^4}(ig)^2\frac{-i}{k^2}\Big(\bar{d}_L\gamma^\nu t^a\frac{i\not{k}}{k^2}\gamma^\mu u_L\Big)\Big(\bar{u}_L\gamma_\mu t^a\frac{i\not{k}}{k^2}\gamma_\nu s_L\Big) \\ &= -ig^2\int\frac{d^4k}{(2\pi)^4}\frac{k^2/d}{(k^2)^3}\big(\bar{d}_L\gamma^\nu t^a\gamma^\lambda\gamma^\mu u_L\big)\big(\bar{u}_L\gamma_\mu t^a\gamma_\lambda\gamma_\nu s_L\big) \\ &= +\frac{g^2}{4}\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^2}\big(\bar{d}_L\gamma^\nu t^a\gamma^\lambda\gamma^\mu u_L\big)\big(\bar{u}_L\gamma_\mu t^a\gamma_\lambda\gamma_\nu s_L\big). \end{aligned} \tag{18.45}$$

The four-fermion operator can be simplified as follows, by the use of the Fierz identity (3.79):

$$\begin{aligned}(\bar{d}_L\gamma^\nu\gamma^\lambda\gamma^\mu u_L)(\bar{u}_L\gamma_\mu\gamma_\lambda\gamma_\nu s_L) &= +(\bar{d}_L\gamma^\nu\gamma^\lambda\gamma^\mu\gamma_\lambda\gamma_\nu s_L)(\bar{u}_L\gamma_\mu u_L)\\ &= +(-2)^2(\bar{d}_L\gamma^\mu s_L)(\bar{u}_L\gamma_\mu u_L)\\ &= 4(\bar{d}_L\gamma^\mu u_L)(\bar{u}_L\gamma_\mu s_L).\end{aligned} \tag{18.46}$$

Again, we must reduce the product of color matrices using identity (18.41), and, again, the first term of this identity will require an additional Fierz transformation. The final result is



$$= +g^2\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^2}\left[\tfrac{1}{2}\bar{u}_L\gamma^\mu u_L\bar{d}_L\gamma_\mu s_L - \tfrac{1}{6}\bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L\right]. \tag{18.47}$$

The last diagram in Fig. 18.3 gives an identical contribution. The sum of the contributions from these four diagrams is

$$-3g^2\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^2}\left[\bar{u}_L\gamma^\mu u_L\bar{d}_L\gamma_\mu s_L - \tfrac{1}{3}\bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L\right]. \tag{18.48}$$

The extraction of the ultraviolet-divergent pieces of the diagrams of Fig. 18.3 is part of our formal prescription for computing the Callan-Symanzik γ function of the weak interaction vertex. However, it is useful to pause at this point and ask about the physical significance of this divergence. The diagrams of Fig. 18.3 would not be divergent if we computed them in the underlying theory with W bosons. In writing the weak interaction as an effective local vertex, we approximated the W boson propagator by a constant, assuming that the momentum k that it carried was much less than m_W:

$$\frac{1}{k^2-m_W^2} \to \frac{-1}{m_W^2}. \tag{18.49}$$

The approximation we used to compute the QCD corrections to the effective vertex is valid only in the region of integration where $k^2 \ll m_W^2$. Outside this region we must use the full W propagator; this introduces an extra factor of k^2 in the denominator and makes the integral converge. Thus, in a direct calculation of the QCD correction, the ultraviolet-divergent terms in the evaluation of Fig. 18.3 would be replaced by logarithms cut off at m_W. The lower limit of the logarithm is set by the external momenta. In the decay of a K meson—the lightest hadron containing the s quark—these are of order m_K. Thus the correction given in (18.48) should be evaluated by replacing

$$g^2\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^2} \to \frac{\alpha_s}{4\pi}\log\frac{m_W^2}{m_K^2}. \tag{18.50}$$

With this interpretation, we can rewrite (18.48) as the order-α_s correction to the leading-order weak interaction vertex. The effect of this correction is

the rescaling and modification of the weak interaction operator:

$$\bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L \rightarrow$$
$$\Big(1+\frac{\alpha_s}{4\pi}\log\frac{m_W^2}{m_K^2}\Big)\bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L - 3\Big(\frac{\alpha_s}{4\pi}\log\frac{m_W^2}{m_K^2}\Big)\bar{u}_L\gamma^\mu u_L\bar{d}_L\gamma_\mu s_L. \tag{18.51}$$

Notice that the QCD corrections not only rescale the normalization of the original operator but also introduce a new operator with a different structure. This calculation makes concrete the idea introduced in Section 12.4 that the diagrams that change the normalization of local operators may also mix together different operators with the same dimension and quantum numbers.

Since the value of the logarithm in (18.50) is about 10, the size of the leading QCD correction is of order 1 and so higher-order corrections are important. To sum the leading logarithmic corrections, we return to the renormalization group analysis. For clarity, define

$$\mathcal{O}^1 = \bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L; \qquad \mathcal{O}^2 = \bar{u}_L\gamma^\mu u_L\bar{d}_L\gamma_\mu s_L. \tag{18.52}$$

We will use the subscript 0 to denote bare operators and the subscript M to denote operators obeying renormalization conditions at the scale M. From the diagrams of Fig. 18.3, we have found that the operator whose matrix elements have the quark structure of $\mathcal{O}^1$, properly normalized at the scale M, is given by

$$\mathcal{O}^1_M = \mathcal{O}^1_0 + \delta^{11}\mathcal{O}^1_0 + \delta^{12}\mathcal{O}^2_0, \tag{18.53}$$

where the δ^{ij} are counterterms,

$$\delta^{11} = -\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}; \qquad \delta^{12} = +3\frac{g^2}{(4\pi)^2}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \tag{18.54}$$

A reciprocal calculation gives $\mathcal{O}^2_M$ in terms of bare operators:

$$\mathcal{O}^2_M = \mathcal{O}^2_0 + \delta^{21}\mathcal{O}^1_0 + \delta^{22}\mathcal{O}^2_0, \tag{18.55}$$

with

$$\delta^{21} = \delta^{12}, \qquad \delta^{22} = \delta^{11}.$$

Then, in the manner than we discussed in Eq. (12.109), the operator rescaling of $\mathcal{O}^1$ and $\mathcal{O}^2$ is described in the Callan-Symanzik equation by a matrix γ^{ij} linking the two operators. Expanding this equation to first order in g^2, we see that this matrix is given to one-loop order by

$$\gamma^{ij} = M\frac{\partial}{\partial M}\big[-\delta^{ij}\big]. \tag{18.56}$$

Thus we find

$$\gamma = \frac{g^2}{(4\pi)^2}\begin{pmatrix} -2 & 6 \\ 6 & -2 \end{pmatrix}, \tag{18.57}$$

acting on the space of operators $\mathcal{O}^1$, $\mathcal{O}^2$.

The simplest way to deduce the physical effects of the rescaling described by (18.57) is to diagonalize this matrix and thus find a new basis of operators that are rescaled without mixing. For the matrix (18.57), the eigenoperators are easily seen to be

$$\begin{aligned} \mathcal{O}^{1/2} &= \tfrac{1}{2}\left[\bar{d}_L\gamma^\mu u_L \bar{u}_L\gamma_\mu s_L - \bar{u}_L\gamma^\mu u_L \bar{d}_L\gamma_\mu s_L\right], \\ \mathcal{O}^{3/2} &= \tfrac{1}{2}\left[\bar{d}_L\gamma^\mu u_L \bar{u}_L\gamma_\mu s_L + \bar{u}_L\gamma^\mu u_L \bar{d}_L\gamma_\mu s_L\right]. \end{aligned} \tag{18.58}$$

The superscripts on these operators are their isospin quantum numbers. The operator $\mathcal{O}^{1/2}$ is antisymmetric under the interchange of the labels $\bar{d}$ and $\bar{u}$; thus, these two isospin-1/2 fields are combined to total isospin zero, and so the whole operator is isospin-1/2. This operator can mediate decays of the K meson that change the isospin by 1/2 unit, such as $K^0 \to \pi^+\pi^-$, but not processes that change the isospin by 3/2, such as $K^+ \to \pi^+\pi^0$. Experimentally, processes of the former type occur almost a thousand times faster (an observation called the $\Delta I = 1/2$ rule). Thus, it is interesting that the hard QCD corrections already make a distinction between these operators.

From the eigenvalues of (18.57), we obtain the Callan-Symanzik γ functions of the eigenoperators (18.58):

$$\gamma_{1/2} = -8\frac{g^2}{(4\pi)^2}; \qquad \gamma_{3/2} = +4\frac{g^2}{(4\pi)^2}. \tag{18.59}$$

According to Eqs. (18.24) and (18.25), this implies that the operator $\mathcal{O}^{1/2}$ receives an enhancement from hard QCD logarithms, while the operator $\mathcal{O}^{3/2}$ receives a suppression. More explicitly, we can write the operator that appears in the original nonleptonic weak interaction vertex (18.31) as

$$\left[\bar{d}_L\gamma^\mu u_L \bar{u}_L\gamma_\mu s_L\right]\Big|_{m_W} = \left[\mathcal{O}^{1/2}\right]\Big|_{m_W} + \left[\mathcal{O}^{3/2}\right]\Big|_{m_W}. \tag{18.60}$$

As above, the subscript refers to the mass scale at which the operator is normalized. We now account for the QCD logarithms associated with evaluating the matrix element of this operator at a lower momentum scale, m_K, by replacing the operators on the right-hand side of (18.60) with operators renormalized at m_K, with the rescaling factor (18.25). This gives

$$\begin{aligned} \left[\bar{d}_L\gamma^\mu u_L \bar{u}_L\gamma_\mu s_L\right]\Big|_{m_W} &= \left(\frac{\log(m_W^2/\Lambda^2)}{\log(m_K^2/\Lambda^2)}\right)^{4/b_0}\left[\mathcal{O}^{1/2}\right]\Big|_{m_K} \\ &\quad + \left(\frac{\log(m_W^2/\Lambda^2)}{\log(m_K^2/\Lambda^2)}\right)^{-2/b_0}\left[\mathcal{O}^{3/2}\right]\Big|_{m_K}, \end{aligned} \tag{18.61}$$

where, again, $b_0 = 11 - \frac{2}{3}n_f$. This equation shows that, unlike the case of semileptonic weak interactions, the overall normalization of the effective Lagrangian for nonleptonic weak interactions is changed by QCD logarithms. In addition, the quark structure of the effective Lagrangian is altered.

Quantitatively, taking $n_f = 4$ and $\Lambda = 150$ MeV as an illustration, we find

$$\left[\bar{d}_L\gamma^\mu u_L\bar{u}_L\gamma_\mu s_L\right]\Big|_{m_W} = 2.1\left[\mathcal{O}^{1/2}\right]\Big|_{m_K} + 0.7\left[\mathcal{O}^{3/2}\right]\Big|_{m_K}. \tag{18.62}$$

Thus, the QCD logarithmic corrections from m_W to m_K give the $\Delta I = 1/2$ part of the effective vertex an enhancement of about a factor of 3.* The observed $\Delta I = 1/2$ rule in K decays requires a factor of 20 enhancement. However, part of this is expected to arise from the ratio of the matrix elements of the operators $\mathcal{O}^{1/2}_{m_K}$ and $\mathcal{O}^{3/2}_{m_K}$ between physical hadron states, which are determined by the soft, nonperturbative part of QCD dynamics.

18.3 The Operator Product Expansion

One way to describe the development of the previous section is to say that we studied an interaction that was fundamentally a product of currents by replacing this product of operators with a single local operator. We then derived the physical consequences of the original, composite, interaction by working out the QCD rescaling of this operator. The procedure of replacing a product of operators with a single effective vertex is useful in many contexts in quantum field theory. Thus, in this section, we will pause from our study of QCD to write out the general formalism governing this procedure.

Let us abstract the situation described in the previous section as follows: Consider a quantum field theory process that includes two operators $\mathcal{O}_1$, $\mathcal{O}_2$ separated by a small distance x, together with other fields $\phi(y_i)$ located much farther away, or together with external physical states. In the example above, the two operators are the quark currents that appear in the weak interaction vertex, and their separation x is a distance of order m_W^{-1}, the range of the W propagator. The external states, which contain K and π mesons, can be described by operators that create and destroy these particles. The amplitude for K decay by the weak interactions, or any more general process of this class, can then be extracted from the Green's function

$$G_{12}(x; y_1, \cdots, y_m) = \langle \mathcal{O}_1(x)\mathcal{O}_2(0)\phi(y_1)\cdots\phi(y_m)\rangle\,, \tag{18.63}$$

considered in the limit $x \to 0$, with the y_i fixed away from the origin. Here and in the following discussion, products of operators will be considered to be time-ordered, just as we would find by writing the product of fields under the functional integral.

The product of operators $\mathcal{O}_1(x)\mathcal{O}_2(0)$ can potentially create the most general local disturbance in the vicinity of the point 0. However, any such disturbance can be described as the effect of a local operator placed at 0. This

*M. K. Gaillard and B. W. Lee, *Phys. Rev. Lett.* **33**, 108 (1974); G. Altarelli and L. Maiani, *Phys. Lett.* **52B**, 351 (1974).

local operator must have the global symmetry quantum numbers of the product of $\mathcal{O}_1\mathcal{O}_2$, but it is otherwise unrestricted. It is useful to write this operator as a linear combination of operators from a standard basis. The coefficients in this linear combination can depend on the separation x. Typically, products of operators in quantum field theory are singular, so it is likely that some of the coefficients will have singularities as $x \to 0$. Combining these observations, Wilson proposed that the effects of the operator product could be computed by replacing the product of operators in (18.63) with a linear combination of local operators,

$$\mathcal{O}_1(x)\mathcal{O}_2(0) \to \sum_n C_{12}{}^n(x)\mathcal{O}_n(0), \tag{18.64}$$

where the coefficients $C_{12}{}^n(x)$ are c-number functions. This *operator product expansion* (OPE) will depend only on the operators $\mathcal{O}_1$, $\mathcal{O}_2$, and their separation and will be independent of the identity and location of the other fields appearing in the Green's function.

The expansion (18.64) implies that the Green's function (18.63) can be expanded for small x as follows:

$$G_{12}(x; y_1, \cdots, y_m) = \sum_n C_{12}{}^n(x) G_n(y_1, \cdots y_m), \tag{18.65}$$

where

$$G_n(y_1, \cdots, y_m) = \langle \mathcal{O}_n(0)\phi(y_1)\cdots\phi(y_m)\rangle, \tag{18.66}$$

and all of the dependence on x is now carried by the OPE coefficient functions. In the example of the previous section, the final amplitudes depended in a rather involved way on the small separation of the two operators, through the dependence of the coefficients in (18.61) on m_W. From the viewpoint of the operator product expansion, this dependence is carried by the coefficient functions and is determined for all matrix elements when these are computed.

In Sections 18.1 and 18.2, we used the renormalization group to compute the enhancement or suppression factors for operator matrix elements. Thus it is natural to expect that the form of the operator product coefficients is also determined by the renormalization group. We will now work out this relation. To begin, we rewrite the expansion (18.64) more precisely. The operators that appear in this relation must be defined at some renormalization scale M. Then the operator product expansion reads:

$$\left[\mathcal{O}_1(x)\right]_M \left[\mathcal{O}_2(0)\right]_M = \sum_n C_{12}{}^n(x; M)\left[\mathcal{O}_n(0)\right]_M. \tag{18.67}$$

Note that the coefficient functions can depend on M, since they must absorb the M-dependent operator rescalings. If we use the left-hand side of (18.67) to compute (18.63), this function obeys the Callan-Symanzik equation

$$\left[M\frac{\partial}{\partial M} + \beta\frac{\partial}{\partial g} + m\gamma + \gamma_1 + \gamma_2\right] G_{12}(x; y_1, \cdots, y_m; M) = 0. \tag{18.68}$$

Similarly, with the operator $\mathcal{O}_n$ normalized at M, the Green's function (18.66) obeys

$$\left[M\frac{\partial}{\partial M}+\beta\frac{\partial}{\partial g}+m\gamma+\gamma_n\right]G_n(y_1,\cdots,y_m;M)=0. \tag{18.69}$$

By applying (18.68) to the right-hand side of (18.65), we see that these relations are consistent only if the OPE coefficient functions obey the Callan-Symanzik equation

$$\left[M\frac{\partial}{\partial M}+\beta\frac{\partial}{\partial g}+\gamma_1+\gamma_2-\gamma_n\right]C_{12}{}^n(x;M)=0. \tag{18.70}$$

We now solve this equation by our standard methods. First, let us apply dimensional analysis. If the operators $\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_n$ have dimensions d_1, d_2, d_n, the coefficient function $C_{12}{}^n(x)$ must have the dimensions of $(\text{mass})^{d_1+d_2-d_n}$. Thus,

$$C_{12}{}^n(x)=\left(\frac{1}{|x|}\right)^{d_1+d_2-d_n}\widetilde{C}(xM), \tag{18.71}$$

where $\widetilde{C}(xM)$ is a dimensionless function. This function is determined from (18.70) according to the method of Section 12.3. Thus,

$$C_{12}{}^n(x)=\left(\frac{1}{|x|}\right)^{d_1+d_2-d_n}c(\bar{g}(1/x))\exp\left[\int_{1/x}^{M}d\log M'\big(\gamma_n-\gamma_1-\gamma_2\big)\right], \tag{18.72}$$

with $c(\bar{g})$ a dimensionless function of the running coupling constant at the separation scale $1/x$.

At a fixed point of the renormalization group, the γ functions would take definite values $\gamma_{j*}=\gamma_j(g_*)$. Then, the solution (18.72) can be evaluated as

$$C_{12}{}^n(x)=\left(\frac{1}{|x|}\right)^{d_1+d_2-d_n}c(g_*)\exp\Big[\log(xM)\big(\gamma_{n*}-\gamma_{1*}-\gamma_{2*}\big)\Big]. \tag{18.73}$$

Thus, in this case,

$$C_{12}{}^n(x)\sim\left(\frac{1}{|x|}\right)^{d_1^*+d_2^*-d_n^*}, \tag{18.74}$$

where

$$d_j^*=d_j+\gamma_j(g_*) \tag{18.75}$$

is the true scaling dimension of the operator $\mathcal{O}_j$ at the fixed point.

For the case of an asymptotically free theory, the scaling relation is complicated in the way that we worked out in Section 18.1. In the leading order of perturbation theory, the three γ functions take the form (18.24). Then the solution of (18.72) takes the form

$$C_{12}{}^n(x)\sim\left(\frac{1}{|x|}\right)^{d_1+d_2-d_n}\left(\frac{\log(1/|x|^2\Lambda^2)}{\log(M^2/\Lambda^2)}\right)^{(a_n-a_1-a_2)/2b_0} \tag{18.76}$$

In the example of Section 18.2, the original operators were currents with dimension 3 and $\gamma = 0$, at separation m_W^{-1}, and the final local operators had dimension 6. Thus, (18.76) does properly reproduce the dependence of (18.61) on m_W. Notice that the renormalization group dependence is less complicated for a product of currents, which have a fixed normalization independent of scale. This special case occurs often in applications of the operator product expansion.

We have written Eq. (18.70) without taking account of operator mixing. However, as we have already seen, operator mixing is often an essential part of the applications of the OPE. It is straightforward to include this effect by rewriting the analysis that leads to (18.70) using matrix-valued γ functions. For example, with operator mixing, the Callan-Symanzik equation for G_n will be modified to

$$\left[\delta_{np}\left(M\frac{\partial}{\partial M} + \beta\frac{\partial}{\partial g} + m\gamma\right) + \gamma_{np}\right] G_p(y_1, \cdots, y_m; M) = 0. \tag{18.77}$$

With these changes, (18.70) becomes

$$\begin{aligned}\left[M\frac{\partial}{\partial M} + \beta\frac{\partial}{\partial g}\right] C_{12}{}^{n}(x;M) + \gamma_{1k} C_{k2}{}^{n}(x;M)& \\ + \gamma_{2k} C_{1k}{}^{n}(x;M) - \gamma_{kn} C_{12}{}^{k}(x;M) &= 0.\end{aligned} \tag{18.78}$$

Notice that the first two γ matrices act on the OPE coefficient from the left, while the third acts from the right. In the case of a product of currents, the first two γ matrices vanish and (18.78) simplifies to

$$\left[M\frac{\partial}{\partial M} + \beta\frac{\partial}{\partial g}\right] C_{12}{}^{n}(x;M) - C_{12}{}^{k}(x;M)\gamma_{kn} = 0. \tag{18.79}$$

This equation will play an important role in the analysis of Section 18.5.

18.4 Operator Analysis of e^+e^- Annihilation

It is not difficult to imagine that there is a connection between matrix elements in which currents are placed at short distances from one another and matrix elements in which currents deliver a hard momentum transfer. Thus we might expect that the idea of the operator product expansion will give us a new viewpoint from which to understand the theory of hard-scattering processes in QCD. In this section and the next, we will work out the relation of the operator product expansion to the perturbative QCD analysis of Chapter 17.

We begin by discussing the total cross section for e^+e^- annihilation to hadrons. Below Eq. (17.9), we argued that this total cross section could be computed in QCD perturbation theory, using a value of α_s corresponding to the scale of the total center of mass energy. However, this argument was a purely intuitive one, with many logical jumps. In this section, we will give a more rigorous argument to the same conclusion.

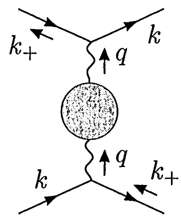


Figure 18.4. Diagrams whose imaginary part yields the total cross section for $e^+e^- \to$ hadrons.

In order to invoke the operator product expansion, we must write the total cross section for e^+e^- annihilation to hadrons as the matrix element of a product of currents. To do this, we use the optical theorem to relate the total e^+e^- scattering cross section to the forward scattering amplitude for $e^+e^- \to e^+e^-$. Ignoring the mass of the electron, we see from Eq. (7.49) that

$$\sigma(e^+e^-) = \frac{1}{2s}\operatorname{Im}\mathcal{M}(e^+e^- \to e^+e^-). \tag{18.80}$$

To compute the cross section for $e^+e^- \to$ hadrons, we consider in the computation of the imaginary part only the contributions from hadronic intermediate states. To leading order in α, but to all orders in the strong interactions, these contributions come from considering only diagrams of the form of Fig. 18.4, and taking the imaginary part of the hadronic contributions to the vacuum polarization.

The value of the diagrams shown in Fig. 18.4 is

$$i\mathcal{M} = (-ie)^2\bar{u}(k)\gamma_\mu v(k_+)\frac{-i}{s}\big(i\Pi_h^{\mu\nu}(q)\big)\frac{-i}{s}\bar{v}(k_+)\gamma_\nu u(k), \tag{18.81}$$

where $s = q^2$ and $\Pi_h^{\mu\nu}(q)$ is the hadronic part of the vacuum polarization. By the Ward identity, this can be written

$$\Pi_h^{\mu\nu}(q) = (q^2 g^{\mu\nu} - q^\mu q^\nu)\Pi_h(q^2). \tag{18.82}$$

The $q^\mu q^\nu$ terms give zero when contracted with the external electron currents, so only the $g^{\mu\nu}$ term survives. To evaluate the electron spinor part of (18.81), we use the fact that, in this forward scattering amplitude, the initial and final momenta and spins are set equal. Then, averaging over the initial spin gives

$$\begin{aligned}\frac{1}{2}\cdot\frac{1}{2}\sum_{\text{spins}}\bar{u}(k)\gamma^\mu v(k_+)\bar{v}(k_+)\gamma_\mu u(k) &= \frac{1}{4}\operatorname{tr}\big[\not{k}\gamma^\mu\not{k}_+\gamma_\mu\big]\\ &= \frac{1}{4}\cdot(-2)\cdot 4(k\cdot\mathbf{k}_+)\\ &= -s.\end{aligned} \tag{18.83}$$

Thus, we find

$$\sigma(e^+e^- \to \text{hadrons}) = -\frac{4\pi\alpha}{s}\operatorname{Im}\Pi_h(s). \tag{18.84}$$

To check this result, we can look back to the one-loop value of Π in QED (7.91), or to the imaginary part of this expression given in Eq. (7.92):

$$\operatorname{Im}\Pi(s+i\epsilon) = -\frac{\alpha}{3}\sqrt{1-\frac{4m^2}{s}}\left(1+\frac{2m^2}{s}\right). \tag{18.85}$$

Combining (18.85) with (18.84), we obtain the correct leading-order cross section for production of a new heavy lepton in e^+e^- annihilation,

$$\sigma(e^+e^- \to L^+L^-) = \frac{4\pi\alpha^2}{3s}\sqrt{1-\frac{4m^2}{s}}\left(1+\frac{2m^2}{s}\right). \tag{18.86}$$

If we multiply (18.86) by a factor of 3 for color and sum over quark flavors with the squares of the quark charges, we obtain the leading-order prediction of QCD.

Now that we have relation (18.84), we complete the connection we wished to prove by noting that the hadronic vacuum polarization is simply a matrix element of a product of currents. Let J^μ be the electromagnetic current of quarks,

$$J^\mu = \sum_f Q_f \bar{q}_f \gamma^\mu q_f. \tag{18.87}$$

Then

$$i\Pi_h^{\mu\nu}(q) = -e^2\int d^4x\, e^{iq\cdot x}\left\langle 0\right| T\{J^\mu(x)J^\nu(0)\}\left|0\right\rangle. \tag{18.88}$$

In the limit in which the point x approaches 0, we can reduce the product of currents by applying the operator product expansion. Since we will be taking the vacuum expectation value of the product, we need only list the contribution from operators that are gauge-invariant Lorentz scalars. Thus,

$$J_\mu(x)J_\nu(0) \sim C_{\mu\nu}{}^{1}(x)\cdot 1 + C_{\mu\nu}{}^{\bar{q}q}(x)\bar{q}q(0) + C_{\mu\nu}{}^{F^2}(x)(F^a_{\alpha\beta})^2(0) + \cdots. \tag{18.89}$$

Note that we have included the operator 1 on the right-hand side, and the next possible operators in QCD have dimension 3 and 4, respectively. Since the operator $\bar{q}q$ violates chiral symmetry, its coefficient function must have an explicit factor of the quark mass. Thus, by dimensional analysis,

$$C_{\mu\nu}{}^{1} \sim x^{-6}, \qquad C_{\mu\nu}{}^{\bar{q}q} \sim mx^{-2}, \qquad C_{\mu\nu}{}^{F^2} \sim x^{-2}, \tag{18.90}$$

and the higher terms in the series are less singular as $x \to 0$.

To compute $\Pi_h^{\mu\nu}(q)$, we need the Fourier transform of the product of currents. Assuming that this Fourier transform is indeed dominated by the limit of short distances, we can compute it by Fourier-transforming the individual OPE coefficients. Since the currents are conserved, the individual terms in the OPE must give zero when dotted with q^μ. Thus the transformed OPE takes

the form

$$-e^2\int d^4x\, e^{iq\cdot x} J^\mu(x)J^\nu(0)$$
$$= -ie^2(q^2g^{\mu\nu} - q^\mu q^\nu)\big[c^1(q^2)\cdot 1 + c^{\bar{q}q}(q^2)\cdot m\bar{q}q + c^{F^2}(q^2)\cdot (F^a_{\alpha\beta})^2 + \cdots\big], \tag{18.91}$$

where the c^i are Lorentz-invariant c-number functions of q^2, and the factor of i at the beginning of the second line is inserted as a convenient convention. By dimensional analysis, we find

$$c^1 \sim (q^2)^0, \qquad c^{\bar{q}q} \sim (q^2)^{-2}, \qquad c^{F^2} \sim (q^2)^{-2}, \tag{18.92}$$

and the higher terms are more irrelevant for large q.

The OPE coefficients $c^i(q^2)$ can be computed from Feynman diagrams. As shown in Fig. 18.5, the coefficient of the operator 1 is the sum of diagrams with no external legs other than the current insertions. The leading QCD diagram is just the simple vacuum polarization diagram, multiplied by the color factor 3 and the sum of the squares of the quark charges. Combining these factors with Eq. (7.91), we have

$$c^1(q^2) = -\Big(3\sum_f Q_f^2\Big)\cdot \frac{\alpha}{3\pi}\log(-q^2). \tag{18.93}$$

The corrections to this result are of order $\alpha_s(q^2)$. The higher coefficient functions are extracted from diagrams with more external legs. For example, the coefficient function of $(F^a_{\alpha\beta})^2$ is determined by diagrams with two external gluon legs.

Still assuming that the Fourier transform of the product of currents can be computed from the OPE for the region of large timelike q^2, we can complete our evaluation of the cross section for $e^+e^- \to$ hadrons by taking the vacuum expectation value of (18.91), extracting the imaginary parts of the coefficient functions, and substituting the result into (18.84). We find

$$\sigma(e^+e^- \to \text{hadrons}) = \frac{4\pi\alpha^2}{s}\big[\mathrm{Im}\, c^1(q^2) + \mathrm{Im}\, c^{\bar{q}q}(q^2)\langle 0|\, m\bar{q}q\,|0\rangle + \mathrm{Im}\, c^{F^2}(q^2)\langle 0|\,(F^a_{\alpha\beta})^2\,|0\rangle + \cdots\big]. \tag{18.94}$$

The first term of this series is just the result of summing perturbative QCD diagrams for the e^+e^- total cross section. The additional terms give corrections to this result which depend on soft hadronic matrix elements, but these corrections are explicitly suppressed at high energy by factors $(q^2)^{-2}$. (Incidentally, this expansion, which applies equally well in the absence of QCD interactions, explains why (18.86) contains no term of order s^{-2} when expanded for large s.) If we insert the leading-order expression (18.93) into (18.94), we obtain the familiar result

$$\sigma(e^+e^- \to \text{hadrons}) = \frac{4\pi\alpha^2}{s}\sum_f Q_f^2. \tag{18.95}$$

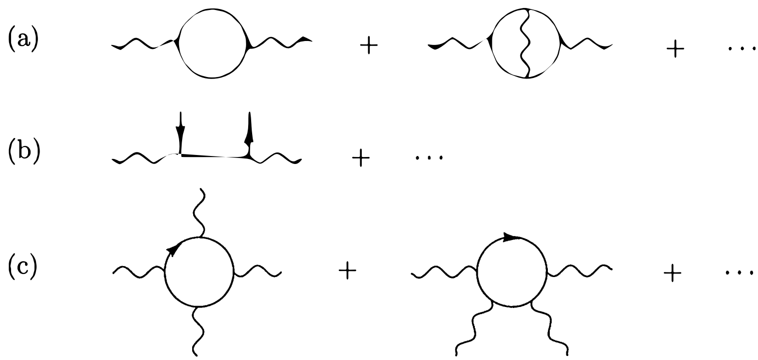


Figure 18.5. Feynman diagrams contributing the operator product coefficient, in the expansion of the product of currents, for the operator (a) 1; (b) $\bar{q}q$; (c) $(F^a_{\alpha\beta})^2$.

Our result (18.94) is pleasing, but the logic that led us to it was not correct. To compute the e^+e^- total cross section, we must compute $\Pi_h(q^2)$ in the region of large *timelike* momentum q, where the expectation value of the product of currents is dominated by intermediate states of high energy, involving large numbers of physical hadrons. Thus we need $\Pi_h(q^2)$ in precisely the region where it is not dominated by short-distance perturbations of the quark and gluon fields. To compute the product of currents from the short-distance expansion, we choose kinematic conditions such that the intermediate states that enter the computation of the product of currents are far off-shell, so that they cannot propagate far from the converging points x and 0. This condition is satisfied at large *spacelike* momentum, or, equivalently, at small spacelike separation. However, it seems at first sight that a computation in this region is useless for determination of the e^+e^- cross section.

Fortunately, there is a wonderful trick for relating the values of a quantum field theory amplitude in two well-separated kinematic regions. This trick, called the method of *dispersion relations*, makes use of the general analytic properties of the amplitude. Since (18.88) is the Fourier transform of a two-point correlation function, we know from the analysis of Section 7.1 that $\Pi_h(q^2)$ possesses a Källén-Lehmann spectral representation. Thus, $\Pi_h(q^2)$ is an analytic function of q^2 with a branch cut on the positive q^2 axis and no other singularities in the complex q^2 plane. This analytic structure is shown in Fig. 18.6. The discontinuity of $\Pi_h(q^2)$ across the branch cut is $(2i)$ times the imaginary part of Π_h and so is directly related to the total e^+e^- annihilation cross section.

With this additional knowledge about $\Pi_h(q^2)$, we can argue as follows. Let $q^2 = -Q_0^2$ be a value sufficiently far into the spacelike region of q that the Fourier transform of the product of currents can be computed from the

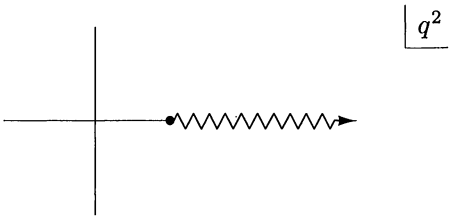


Figure 18.6. Analytic singularities of $\Pi_h(q^2)$ in the complex q^2 plane.

operator product expansion. Now consider the integral

$$I_n = -4\pi\alpha \oint \frac{dq^2}{2\pi i} \frac{1}{(q^2+Q_0^2)^{n+1}} \Pi_h(q^2), \tag{18.96}$$

for $n \geq 1$, evaluated on a contour encircling $q^2 = -Q_0^2$. If we contract the contour onto the pole, we find

$$I_n = -4\pi\alpha \frac{1}{n!} \frac{d^n}{d(q^2)^n} \Pi_h \bigg|_{q^2=-Q_0^2}, \tag{18.97}$$

which can be computed by evaluating Π_h from the the operator product relation (18.91),

$$\Pi_h(q^2) = -e^2 \big[c^1(q^2) + c^{\bar{q}q}(q^2) \langle 0| m\bar{q}q |0\rangle + c^{F^2}(q^2) \langle 0| (F^a_{\alpha\beta})^2 |0\rangle + \cdots \big]. \tag{18.98}$$

On the other hand, we can evaluate the integral by distorting the contour to the form of Fig. 18.7. Since none of the coefficient functions grow faster than $(q^2)^0$ times logarithms as $q^2 \to \infty$, the contour at infinity can be neglected for $n \geq 1$. The piece of the contour that wraps around the branch cut gives

$$\begin{aligned} I_n &= -4\pi\alpha \int \frac{dq^2}{2\pi i} \frac{1}{(q^2+Q_0^2)^{n+1}} \operatorname{Disc} \Pi_h(q^2) \\ &= -4\pi\alpha \int \frac{dq^2}{2\pi} \frac{1}{(q^2+Q_0^2)^{n+1}} \frac{1}{i} 2i \operatorname{Im} \Pi_h(q^2) \\ &= \frac{1}{\pi} \int_0^\infty ds \, \frac{s}{(s+Q_0^2)^{n+1}} \sigma(s). \end{aligned} \tag{18.99}$$

This is an integral over the total cross section for $e^+e^- \to$ hadrons. By equating (18.97) and (18.99), we obtain a series of integral relations between the OPE coefficients, evaluated in QCD perturbation theory, and the observable cross section. These relations, which were first constructed by Novikov, Shifman, Voloshin, Vainshtein, and Zakharov, are known as the *ITEP sum rules.*†

†The theory of these sum rules is reviewed in V. A. Novikov, L. B. Okun, M. A. Shifman, A. I. Vainshtein, M. B. Voloshin, and V. I. Zakharov, *Phys. Repts.* **41**, 1 (1978).

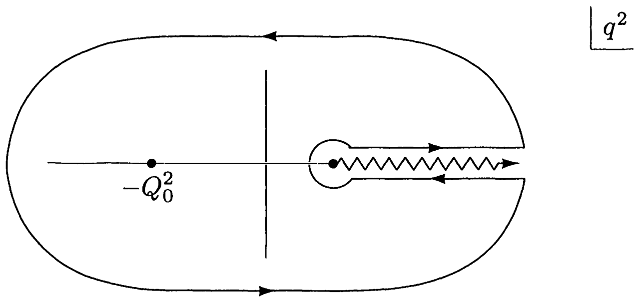


Figure 18.7. Contour of integration involved in the derivation of the ITEP sum rules for $\sigma(e^+e^- \to \text{hadrons})$.

Evaluating the sum rules with only the leading QCD expression for $c^1(q^2)$, we find

$$\int_0^\infty ds \frac{s}{(s+Q_0^2)^{n+1}}\sigma(s) = \frac{4\pi\alpha^2}{n(Q_0^2)^n}\sum_f Q_f^2 + \mathcal{O}(\alpha_s(Q_0^2)) + \mathcal{O}((Q_0^2)^{-2}). \quad (18.100)$$

The leading-order relation is consistent with the lowest-order cross section given in Eq. (18.95). The corrections come from higher orders of QCD perturbation theory, with α_s taken at the scale Q_0^2, and from the higher operator terms in the OPE.

If the correction terms in (18.100) converged to zero uniformly in n, we could invert the sum rules and derive from them our result (18.94). However, the true situation is more subtle. Because the derivatives in (18.97) emphasize terms with stronger q^2 variation, the correction terms in the ITEP sum rules are more and more important as n increases. Thus the most important deviations of the cross section from the prediction of QCD perturbation theory are oscillations about this prediction, which average out in the sum rules for low n. The comparison of theory and experiment is shown in Fig. 18.8. At large s, (18.94) is quite accurate. As s becomes smaller, however, the oscillations grow in size. Eventually, they come to dominate the total cross section as the resonances associated with quark-antiquark bound states.

18.5 Operator Analysis of Deep Inelastic Scattering

We now apply the operator product expansion to another example of a QCD hard-scattering process, deep inelastic electron scattering. In Chapter 17 we found that the predictions of QCD for deep inelastic scattering are precise but also intricate in structure. At a first level, QCD implies that deep inelastic scattering is described by the parton model, in which the incident electron scatters from quarks and antiquarks that carry fractions of the total momentum of the proton. These fractions are determined by parton distribution

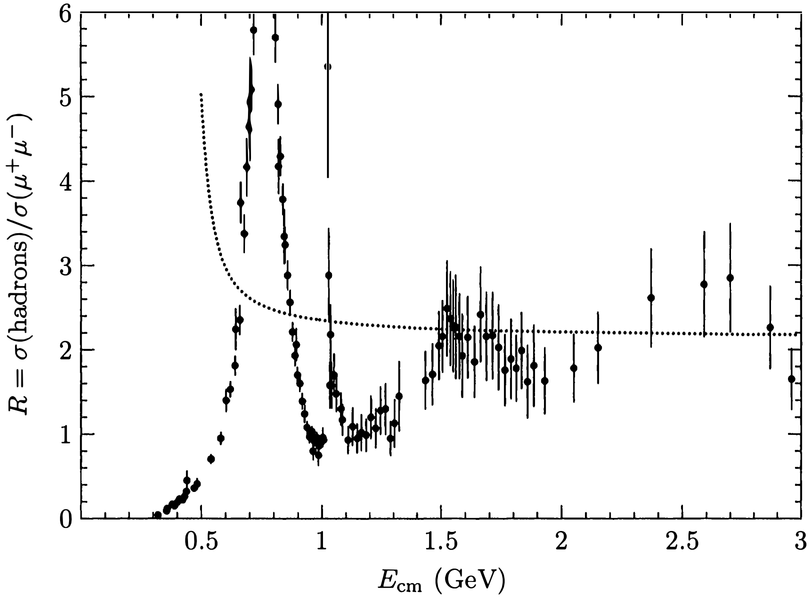


Figure 18.8. Experimental measurements of the total cross section for the reaction $e^+e^- \to$ hadrons at energies below 3 GeV, compared to the prediction of perturbative QCD for 3 quark flavors. The data are taken from the compilation of M. Swartz, *Phys. Rev.* **D**53, 5268 (1996). Complete references to the various results are given there.

functions, which reflect the form of the proton wavefunction and are determined by soft QCD dynamics. However, we saw in Section 17.5 that effects of QCD perturbation theory cause the parton distributions to change their form as a function of the momentum transfer Q^2. We will now show that much of this picture can be reconstructed from our new viewpoint, using the operator product expansion.

In the previous section, we derived the OPE relations for the e^+e^- annihilation cross section in three steps. First, we used the optical theorem to relate this cross section to a matrix element of a product of currents. Second, we applied the operator product expansion to the product of currents. Unfortunately, this expansion could be used only in an unphysical kinematic region. However, in the third step, we used the method of dispersion relations to connect this unphysical result to an integral over the cross section we wished to predict. In our discussion of deep inelastic scattering, we will go through these same three steps. To obtain our final result, we will need to add a fourth step, involving QCD operator rescaling.

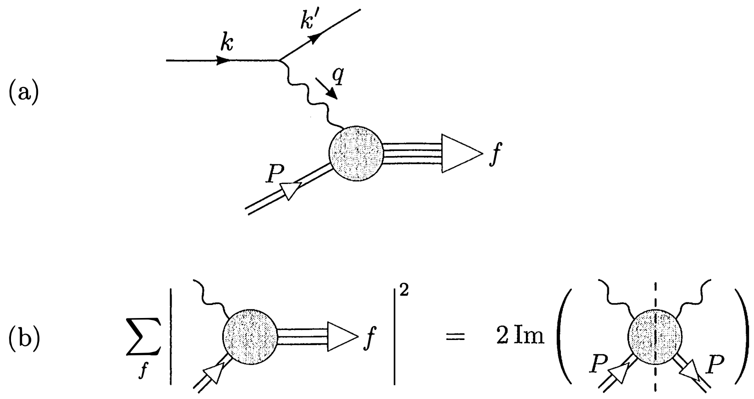


Figure 18.9. Computation of the cross section for deep inelastic electron scattering: (a) general structure of the amplitudes; (b) application of the optical theorem.

Kinematics of Deep Inelastic Scattering

We begin by writing a general expression for the deep inelastic scattering cross section. The matrix element for deep inelastic electron scattering to a final state f is computed as shown in Fig. 18.9(a):

$$i\mathcal{M}(ep \to ef) = (-ie)\bar{u}(k')\gamma_\mu u(k)\frac{-i}{q^2}(ie)\int d^4x\, e^{iq\cdot x}\,\langle f|\, J^\mu(x)\,|P\rangle\,, \quad (18.101)$$

where $J^\mu(x)$ is the quark electromagnetic current (18.87). The core of this expression is the hadronic matrix element of the current between the proton and some high-energy hadronic state. This matrix element must be squared and summed over possible final states. That sum can be computed, using the optical theorem, by relating it to the forward matrix element of two currents in the proton state, as shown in Fig. 18.9(b). Define

$$W^{\mu\nu} = i\int d^4x\, e^{iq\cdot x}\,\langle P|\, T\{J^\mu(x)J^\nu(0)\}\,|P\rangle\,, \quad (18.102)$$

averaged over the spin of the proton. This object is known as the *forward Compton amplitude*, since if it is evaluated at $q^2 = 0$ and contracted with physical polarization vectors, it gives the forward amplitude for photon-proton scattering:

$$i\mathcal{M}(\gamma p \to \gamma p) = (ie)^2\epsilon^*_\mu(q)\epsilon_\nu(q)\big(-iW^{\mu\nu}(P,q)\big). \quad (18.103)$$

However, in the following discussion we will need to analyze (18.102) for general spacelike q and for general polarization states.

The optical theorem for Compton scattering from a proton is

$$2\,\mathrm{Im}\,\mathcal{M}(\gamma p \to \gamma p) = \sum_f \int d\Pi_f \big|\mathcal{M}(\gamma p \to f)\big|^2. \quad (18.104)$$

In the generalization given in (7.49), this result extends to the more general situation in which the initial and final photon polarizations can differ arbitrarily. Transcribing (18.104) to $W^{\mu\nu}$, we find

$$2\,\mathrm{Im}\,W^{\mu\nu}(P,q) = \sum_f \int d\Pi_f \, \langle P|\, J^\mu(-q)\,|f\rangle \, \langle f|\, J^\nu(q)\,|P\rangle \,, \tag{18.105}$$

where $J^\mu(q)$ is the Fourier transform of the current.

We can now compute the deep inelastic cross section in terms of $W^{\mu\nu}$, using (18.105) to represent the square of the last factor. The cross section should be averaged over initial and summed over final electron spins. Thus,

$$\begin{aligned}\sigma(ep \to eX) = \frac{1}{2s}\int \frac{d^3k'}{(2\pi)^3}\frac{1}{2k'}e^4\frac{1}{2}\sum_{\text{spins}}\big[\bar{u}(k)\gamma_\mu u(k')\bar{u}(k')\gamma_\nu u(k)\big] \\ \cdot \Big(\frac{1}{Q^2}\Big)^2 \cdot 2\,\mathrm{Im}\,W^{\mu\nu}(P,q).\end{aligned} \tag{18.106}$$

The electron spinor product can be evaluated as

$$\begin{aligned}\tfrac{1}{2}\sum_{\text{spins}}\big[\bar{u}(k)\gamma_\mu u(k')\bar{u}(k')\gamma_\nu u(k)\big] &= \tfrac{1}{2}\,\mathrm{tr}\big[\not{k}\gamma_\mu \not{k}'\gamma_\nu\big] \\ &= 2\big(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu} k\cdot k'\big).\end{aligned} \tag{18.107}$$

It is useful to convert the integral over the final electron momentum k' and scattering angle θ to an integral over the dimensionless variables x and y that we introduced in Section 17.3. These variables are given in terms of the initial and final electron energies k and k' by

$$x = \frac{Q^2}{2P\cdot q} = \frac{2kk'(1-\cos\theta)}{2m(k-k')}, \qquad y = \frac{2P\cdot q}{2P\cdot k} = \frac{k-k'}{k}. \tag{18.108}$$

Then

$$\left|\frac{\partial(x,y)}{\partial(k',\cos\theta)}\right| = \frac{2k'}{2m(k-k')} = \frac{2k'}{ys}, \tag{18.109}$$

and so

$$\int \frac{d^3k'}{(2\pi)^3}\frac{1}{2k'} = \int \frac{2\pi dk' k' d\cos\theta}{(2\pi)^3\cdot 2} = \int dx dy \frac{ys}{(4\pi)^2}. \tag{18.110}$$

Using (18.107) and (18.110) to simplify (18.106), we find

$$\frac{d^2\sigma}{dxdy}(ep\to eX) = \frac{2\alpha^2 y}{(Q^2)^2}\big(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu}k\cdot k'\big)\,\mathrm{Im}\,W^{\mu\nu}(P,q). \tag{18.111}$$

To go further, we need to know something about the structure of $W^{\mu\nu}$. In the previous section, we used current conservation to write the matrix element of currents in terms of a single scalar function $\Pi_h(q^2)$, as in Eq. (18.82). In

the case of the forward Compton amplitude, the Ward identity again requires

$$q_\mu W^{\mu\nu} = q_\nu W^{\mu\nu} = 0, \tag{18.112}$$

but now there are two possible tensors built from P and q that satisfy these constraints. Thus the forward Compton amplitude is written as an expression involving two scalar form factors:

$$W^{\mu\nu} = \Big(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\Big)W_1 + \Big(P^\mu - q^\mu \frac{P\cdot q}{q^2}\Big)\Big(P^\nu - q^\nu \frac{P\cdot q}{q^2}\Big)W_2. \tag{18.113}$$

The scalar functions W_1, W_2 depend on the two invariants of the problem, $(P\cdot q)$ and q^2, or, alternatively, x and Q^2. If we insert (18.113) into (18.111) and use the fact that dotting q^μ with the lepton tensor gives zero, we find

$$\begin{aligned} \frac{d^2\sigma}{dxdy}(ep\to eX) &= \frac{2\alpha^2 y}{(Q^2)^2}\big[2k\cdot Pk'\cdot P\,\mathrm{Im}\,W_2 + 2k\cdot k'\,\mathrm{Im}\,W_1\big] \\ &= \frac{\alpha^2 y}{Q^4}\big[s^2(1-y)\,\mathrm{Im}\,W_2 + 2xys\,\mathrm{Im}\,W_1\big]. \end{aligned} \tag{18.114}$$

Expression (18.114) is completely general and makes no assumptions about the nature of the strong interactions. It is also rather formal. However, we can easily get an idea of the relation of this formula to our earlier analysis by evaluating $W^{\mu\nu}$ in the parton model and working out the parton expressions for W_1 and W_2. In the parton model, we replace the proton matrix element in (18.102) by a sum of quark matrix elements, weighted with the parton distribution functions. Thus,

$$W^{\mu\nu} \approx i\int d^4x\, e^{iq\cdot x}\int_0^1 d\xi \sum_f f_f(\xi)\cdot\frac{1}{\xi}\left\langle q_f(p)\right| T\big\{J^\mu(x)J^\nu(0)\big\}\left|q_f(p)\right\rangle\Big|_{p=\xi P}. \tag{18.115}$$

The factor $(1/\xi)$ in front of the matrix element gives the proper normalization of the proton state in terms of the quark states. The simplest way to understand this factor is to note that the kinematic prefactor $(1/2s)$ in (18.106) and in other expressions involving an initial-state proton becomes $(1/2\xi s)$, under the ξ integral, in the parton model.

We now evaluate the matrix element in (18.115) using noninteracting fermions. There are two Feynman diagrams, shown in Fig. 18.10. The first diagram on the right in Fig. 18.10 has the value

$$i\int_0^1 d\xi \sum_f f_f(\xi)\frac{1}{\xi}Q_f^2\bar{u}(p)\gamma^\mu \frac{i(\not{p}+\not{q})}{(p+q)^2+i\epsilon}\gamma^\nu u(p); \tag{18.116}$$

the second diagram gives a contribution identical to this one after the interchange of q, μ with $(-q)$, ν. To evaluate (18.116), we average over the quark

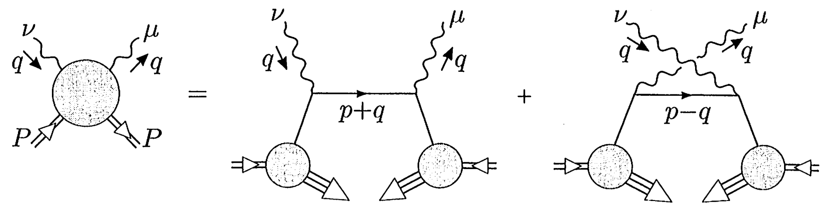


Figure 18.10. Evaluation of $W^{\mu\nu}$ in the parton model.

spin to find

$$\int_0^1 d\xi \sum_f f_f(\xi)\frac{1}{\xi}\cdot\frac{1}{2}\operatorname{tr}\left[\not p\gamma^\mu(\not p+\not q)\gamma^\nu\right]\frac{-1}{2p\cdot q+q^2+i\epsilon}$$

$$= \int_0^1 d\xi \sum_f f_f(\xi)\frac{1}{\xi}\cdot 2\left(p^\mu(p+q)^\nu + p^\nu(p+q)^\mu - g^{\mu\nu}p\cdot(p+q)\right)$$

$$\cdot\frac{-1}{2\xi P\cdot q - Q^2 + i\epsilon}. \tag{18.117}$$

The imaginary part of this expression, which we need to evaluate (18.114), comes from the last factor in (18.117):

$$\operatorname{Im}\left(\frac{-1}{2\xi P\cdot q - Q^2 + i\epsilon}\right) = \pi\delta(2\xi P\cdot q - Q^2) = \frac{\pi}{ys}\delta(\xi - x). \tag{18.118}$$

In the second diagram of Fig. 18.10, the two factors in the denominator have a relative + sign, so this diagram has no imaginary part in the physical region for deep inelastic scattering. Thus, we find that in the parton model,

$$\operatorname{Im} W^{\mu\nu} = \sum Q_f^2 f_f(x)\frac{1}{x}\frac{\pi}{ys}\left(4x^2P^\mu P^\nu + 2x(P^\mu q^\nu + P^\nu q^\mu) - g^{\mu\nu}xys\right). \tag{18.119}$$

By adding and subtracting terms proportional to $q^\mu q^\nu$, we can see that this expression is of the form (18.113), with

$$\operatorname{Im} W_1 = \pi\sum_f Q_f^2 f_f(x), \qquad \operatorname{Im} W_2 = \frac{4\pi}{ys}\sum_f Q_f^2 x f_f(x). \tag{18.120}$$

The parton model expressions for W_1 and W_2 obey the relation

$$\operatorname{Im} W_1 = \frac{ys}{4x}\operatorname{Im} W_2. \tag{18.121}$$

This is another form of the Callan-Gross relation, since the substitution of (18.121) into (18.114) gives

$$\frac{d^2\sigma}{dxdy}(ep\to eX) = \frac{\alpha^2 ys^2}{2Q^4}\left[1+(1-y)^2\right]\operatorname{Im} W_2, \tag{18.122}$$

with the y dependence characteristic of free fermions, as in Eq. (17.125). Finally, substituting from (18.120) for the imaginary part of W_2, we recover this parton model expression precisely:

$$\frac{d^2\sigma}{dxdy}(ep \to eX) = \frac{2\pi\alpha^2 s}{Q^4}\Big(\sum_f Q_f^2 x f_f(x)\Big)\big[1+(1-y)^2\big]. \tag{18.123}$$

This equation will give us a reference point for comparison with more general expressions that we will derive as we continue our analysis.

Expansion of the Operator Product

Since the forward Compton amplitude is a matrix element of a product of currents, an alternative strategy for calculating $W^{\mu\nu}$ is to expand this product as a series of local operators. Like the parton model evaluation, this method makes use of asymptotic freedom. However, in this case, the assumption is applied more directly. The computation of the operator product coefficients will take place explicitly at a small distance of order $1/Q$, and so we can calculate these coefficients in a perturbation theory whose coupling constant is $\alpha_s(Q^2)$.

In the previous section, we computed the coefficients of operators that contribute to the vacuum expectation value of the product of currents by considering the various ways of contracting the quark fields in the product. Here, we should note that the operator 1 does not contribute to the Compton scattering amplitude. The leading contributions come from operators that can create and annihilate quarks in the proton wavefunction.

The most important terms in the operator product of two currents J^μ come from products of two quark currents $\bar{q}_f\gamma^\mu q_f$ with quarks of the same flavor. Therefore we will begin by studying the OPE of the individual quark currents. To zeroth order in α_s, the leading terms of the operator product of quark currents are given by

$$\begin{aligned}&\bar{q}\gamma^\mu q(x)\,\bar{q}\gamma^\nu q(0)\\ &\quad = \bar{q}(x)\gamma^\mu \overbracket{q(x)\bar{q}(0)}\gamma^\nu q(0) + \overbracket{\bar{q}(x)\gamma^\mu q(x)\bar{q}(0)\gamma^\nu q}(0) + \cdots,\end{aligned} \tag{18.124}$$

where the contractions should be evaluated as Feynman propagators for the quark fields. The terms with explicit contractions are singular as $x \to 0$; the remaining terms are nonsingular and thus less important in the short-distance limit. In the OPE of currents with quarks of different flavor, there are no corresponding singular terms; we will argue below that this conclusion is valid even beyond the leading order in α_s.

To evaluate $W^{\mu\nu}$, we must take the Fourier transform of the terms in (18.124), as indicated in (18.102). When we do this, we should remember that the propagators carry not only the Fourier transform momentum q but also whatever momentum is carried in through the quark fields. To take account

of this, it is convenient to represent

$$\int d^4x\, e^{iq\cdot x}\, \bar{q}(x)\gamma^\mu \overbrace{q(x)\bar{q}}(0)\gamma^\nu q(0) = \bar{q}\gamma^\mu \frac{i(i\partial\!\!\!/ + q\!\!\!/)}{(i\partial + q)^2}\gamma^\nu q(0), \tag{18.125}$$

where the derivatives ∂ act to the right on the quark field. Notice that this contribution has the structure of the first diagram on the right in Fig. 18.10. Similarly, the second contraction indicated in (18.124) has the form of the second diagram in Fig. 18.10.

In the short-distance limit, the momentum q will be larger than any external momentum entering the quark fields. Thus we should expand

$$\frac{1}{(i\partial + q)^2} = \frac{-1}{Q^2 - 2iq\cdot\partial + \partial^2} = -\frac{1}{Q^2}\sum_{n=0}^{\infty}\Big(\frac{2iq\cdot\partial - \partial^2}{Q^2}\Big)^n. \tag{18.126}$$

We will argue below that the terms with ∂^2 in the numerator are unimportant and may be dropped. However, we should retain all powers of the ratio $(2iq\cdot\partial/Q^2)$. This ratio has Q^2 in the denominator and so is formally suppressed in the short-distance limit. However, in the parton model

$$\frac{2iq\cdot\partial}{Q^2} \to \frac{2q\cdot\xi P}{Q^2} = 1, \tag{18.127}$$

so, eventually, all of these terms must be equally important. We will see how this works in a moment.

The last step required to reduce the operator product (18.124) to a useful form is to reduce the product of Dirac matrices. We know from (18.113) that, after we average over the proton spin, $W^{\mu\nu}$ will be symmetric under the interchange of μ and ν. Thus, it does no harm to symmetrize the OPE. We can then reduce the product of three Dirac matrices to one by using the identity

$$\tfrac{1}{2}\big(\gamma^\mu\gamma^\alpha\gamma^\nu + \gamma^\nu\gamma^\alpha\gamma^\mu\big) = g^{\mu\alpha}\gamma^\nu + \gamma^\mu g^{\alpha\nu} - g^{\mu\nu}\gamma^\alpha, \tag{18.128}$$

which is easily proved from the anticommutation relations. By the use of (18.126) and (18.128), we can rewrite (18.125) as

$$-i\bar{q}\Big(\gamma^\mu(i\partial^\nu) + \gamma^\nu(i\partial^\mu) - ig^{\mu\nu}\partial\!\!\!/ + \gamma^\mu q^\nu + \gamma^\nu q^\mu - g^{\mu\nu}q\!\!\!/\Big)\frac{1}{Q^2}\sum_{n=0}^{\infty}\Big(\frac{2iq\cdot\partial}{Q^2}\Big)^n q. \tag{18.129}$$

We can remove the term $(i\partial\!\!\!/)q$, which vanishes to leading order in α_s, since the quark field obeys the Dirac equation. To compute W_1 and W_2, we can also drop the terms with explicit factors of q^μ, since these will eventually be organized into the general form (18.113). Then, finally, (18.125) takes the form

$$\begin{aligned}\int d^4x\, e^{iq\cdot x}&\bar{q}(x)\gamma^\mu \overbrace{q(x)\bar{q}}(0)\gamma^\nu q(0)\\ &= -i\bar{q}\big(2\gamma^\mu(i\partial^\nu) - g^{\mu\nu}q\!\!\!/\big)\frac{1}{Q^2}\sum_{n=0}^{\infty}\Big(\frac{2iq\cdot\partial}{Q^2}\Big)^n q,\end{aligned} \tag{18.130}$$

symmetrized under $\mu \leftrightarrow \nu$.

The second term in (18.124) differs from the first by the interchange of the points x and 0 and the interchange of indices μ and ν. Its Fourier transform is thus given by (18.130) with the replacement $q \to -q$. The complete operator product therefore contains only terms even in q. All remaining contributions from the singular terms of the operator product contain the operator

$$\bar{q}\gamma^{\mu_1}(i\partial^{\mu_2})\cdots(i\partial^{\mu_k})q, \tag{18.131}$$

with an even number of indices, with these indices either identified with μ or ν or contracted with powers of q. To write the relevant terms of the operator product expansion, we will modify this operator in two ways. First, since the operator in (18.131) has n vector indices, it contains components that transform under many different irreducible representations of the Lorentz group. Each component has a different rescaling law under renormalization. However, we will see below that only the component of (18.131) with the highest spin is relevant to our analysis. This component is obtained by totally symmetrizing the indices $\mu_1, \ldots, \mu_n$ and then subtracting terms proportional to $g^{\mu_i\mu_j}$ so that the operator is traceless on all pairs of indices. We will retain only this component when we write out the operator product of currents. Second, the operator (18.131) does not transform simply under gauge transformations. Since the original currents J^μ were invariant to color gauge transformations, the operator product of two currents must be a sum of gauge-invariant operators. We can make (18.131) gauge-invariant by replacing each factor of $(i\partial^\mu)$ with a covariant derivative (iD^μ). This modification adds only terms proportional to the strong coupling constant g, so it has no effect on our derivation of the operator product coefficients.

Incorporating these changes, let us define a spin-n operator with quarks of flavor f as follows:

$$\mathcal{O}_f^{(n)\mu_1\cdots\mu_n} = \bar{q}_f\gamma^{\{\mu_1}(iD^{\mu_2})\cdots(iD^{\mu_n\}})q_f - \text{traces}, \tag{18.132}$$

with indices symmetrized and with appropriate subtractions. We can use these operators to write a final expression for the most singular part of the OPE of two currents J^μ. The leading terms in this operator product come from (18.130) and the corresponding contraction with $q \leftrightarrow -q$. Extracting the pieces of these expressions that contain the highest spin operators (18.132), we find

$$\begin{aligned} i\int d^4x\, e^{iq\cdot x} J^\mu(x)J^\nu(0) \\ = \sum_f Q_f^2 \bigg[4\sum_{n=2}^{\infty} \frac{(2q^{\mu_1})\cdots(2q^{\mu_{n-2}})}{(Q^2)^{n-1}} \mathcal{O}_f^{(n)\mu\nu\mu_1\cdots\mu_{n-2}} \\ - g^{\mu\nu}\sum_{n=2}^{\infty} \frac{(2q^{\mu_1})\cdots(2q^{\mu_n})}{(Q^2)^n} \mathcal{O}_f^{(n)\mu_1\cdots\mu_n} \bigg] + \cdots, \end{aligned} \tag{18.133}$$

where the sums over n run over even integers only.

Expression (18.133) has been derived in the leading order in α_s. Higher-order Feynman diagrams will contribute corrections to the coefficient functions of order $\alpha_s(Q^2)$. These corrections will be important only if they are multiplied by large logarithms. If we consider the operators $\mathcal{O}_f^{(n)}$ appearing on the right-hand side to be normalized at the renormalization scale Q, there is no large ratio of momenta available to enhance the QCD corrections to the coefficient functions. Large logarithmic corrections may still arise at a later stage of the calculation, when we compute the matrix elements of the operators $\mathcal{O}_f^{(n)}$.

From the expansion (18.133), it is straightforward to compute an expansion for $W^{\mu\nu}$ by taking its expectation value in the proton state. To carry out this computation, we need to know the proton matrix elements of the operators $\mathcal{O}_f^{(n)}$. Notice that these matrix elements cannot depend on the direction of the momentum q^μ, since that dependence has been isolated in the coefficient functions. This means that only the proton momentum P^μ is available to carry the vector indices of the matrix element. We can therefore write the spin-averaged matrix element of $\mathcal{O}_f^{(n)}$ as

$$\langle P| \, \mathcal{O}_f^{(n)\mu_1\cdots\mu_n} \, |P\rangle = A_f^n \cdot 2P^{\mu_1}\cdots P^{\mu_n} - \text{traces}. \tag{18.134}$$

The coefficients A_f^n are dimensionless. They are not quite pure numbers, because they depend on the renormalization scale of the operators, but we will treat them as constants in the next few paragraphs.

For the case $n=1$, the operators $\mathcal{O}_f^{(1)}$ reduce simply to the quark flavor currents $\bar{q}\gamma^\mu q$; in this case the operators are normalized independently of any scale and the coefficients A_f^1 are truly constants. From our general discussion of form factors in Section 6.2, we know that the proton matrix element of a conserved flavor current at zero momentum transfer is given by

$$\langle P| \, \bar{q}_f\gamma^\mu q_f \, |P\rangle = \bar{u}(P)\gamma^\mu u(P) F_{f1}(0), \tag{18.135}$$

where $F_{f1}(0)$ is equal to the value of the corresponding conserved charge in the proton state. For the quark currents, this charge is just the number of quarks (minus antiquarks) of flavor f in the state $|P\rangle$, which we will call N_f. Averaging (18.135) over the proton spin, we find

$$\langle P| \, \bar{q}_f\gamma^\mu q_f \, |P\rangle = 2P^\mu \cdot N_f. \tag{18.136}$$

Thus, for $n=1$,

$$A_f^1 = N_f = \begin{cases} 2 & f=u, \\ 1 & f=d. \end{cases} \tag{18.137}$$

Similarly, $\mathcal{O}_f^{(2)}$ is the contribution of the quark flavor f to the energy-momentum tensor of QCD:

$$\left(T^{\mu\nu}\right)_f = \bar{q}_f\gamma^{\{\mu}(iD^{\nu\}})q_f. \tag{18.138}$$

Thus, A_f^2 is the fraction of the total energy-momentum of the proton that is carried by the quark flavor f.

When we evaluate the series for $W^{\mu\nu}$ using (18.133) and the expression (18.134) for the operator matrix elements, we find

$$W^{\mu\nu} = \sum_f Q_f^2 \left[8 \sum_n P^\mu P^\nu \frac{(2q\cdot P)^{n-2}}{(Q^2)^{n-1}} A_f^n - 2g^{\mu\nu} \sum_n \frac{(2q\cdot P)^n}{(Q^2)^n} A_f^n \right] + \cdots, \tag{18.139}$$

where the sums over n run over even integers from 2 to infinity. In addition to the corrections to the OPE omitted in (18.133), we have also dropped contributions from the trace terms in (18.134). This is quite appropriate: In each of these terms, two factors of the proton momentum $P^\alpha P^\beta$ are replaced by $g^{\alpha\beta} m_p^2$, were $m_p^2 = P^2$ is the proton mass. When the indices are contracted with powers of q, we obtain a term of order

$$m_p^2 Q^2 \ll (2q\cdot P)^2. \tag{18.140}$$

Since $(Q^2/2P\cdot q) = x$, which is held fixed in deep inelastic scattering as Q^2 becomes large, the contribution from the trace terms is suppressed by a factor m_p^2/Q^2, times powers of x.

In general, an operator of dimension d has a coefficient function in the operator product expansion of currents that has dimension $(\text{mass})^{6-d}$; in the Fourier transform of the OPE, this coefficient function will carry a suppression factor

$$\left(\frac{1}{Q}\right)^{d-2}. \tag{18.141}$$

However, if the operator has spin s, the operator matrix element will contribute s factors of the vector P^μ, so that, in the kinematic region of deep inelastic scattering, the contribution will be of order

$$\left(\frac{2P\cdot q}{Q^2}\right)^s \left(\frac{1}{Q}\right)^{d-s-2}. \tag{18.142}$$

Thus, the relative size of contributions from the OPE to deep inelastic scattering is controlled, not exactly by the dimension of the operator, but rather by the *twist*, defined as

$$t = d - s. \tag{18.143}$$

In our selection of the leading terms in the operator product expansion of currents, we have consistently kept the contribution of leading spin for each dimension or for each power of Q^{-1} in the coefficient. The operators $\mathcal{O}_f^{(n)}$ all have twist $t = 2$, which is the smallest possible value for QCD operators other than the operator 1.

In the operator product of two different flavor currents—for example, $\bar{u}\gamma^\mu u$ and $\bar{d}\gamma^\nu d$—the leading terms in the OPE have the quark structure $(\bar{u}\Gamma u \bar{d}\Gamma d)$ and thus have twist $t \geq 4$. Thus, to all orders in α_s, the cross terms in the operator product of currents J^μ are suppressed by at least a factor

$(1/Q^2)$ relative to the leading-twist terms presented in (18.133). If we neglect these suppressed terms, the expression for $W^{\mu\nu}$ separates, to all orders, into a sum of contributions

$$W^{\mu\nu} = \sum_f Q_f^2 W_f^{\mu\nu}, \tag{18.144}$$

where $W_f^{\mu\nu}$ is the matrix element of two quark flavor currents $\bar{q}_f \gamma^\mu q_f$.

We can read from (18.139) the following expressions for W_1 and W_2:

$$\begin{aligned} W_1 &= \sum_f Q_f^2 \sum_n 2 \frac{(2q \cdot P)^n}{(Q^2)^n} A_f^n, \\ W_2 &= \sum_f Q_f^2 \sum_n \frac{8}{Q^2} \frac{(2q \cdot P)^{n-2}}{(Q^2)^{n-2}} A_f^n, \end{aligned} \tag{18.145}$$

where the sum over n in each line runs over even integers from 2 to infinity. Like (18.139), these expressions explicitly separate according to (18.144). It is noteworthy that the series (18.145) satisfy the Callan-Gross relation in the form (18.121), without further parton model input. However, this relation is corrected in order α_s due to the next-order contributions to the operator product coefficients.

Because the leading contributions to the deep inelastic form factors can be written as sums over quark flavors, it is tempting to reverse the logic of Eq. (18.120) and use these equations to define the parton distribution functions. In particular, let us define

$$x f_f^+(x, Q^2) = \frac{ys}{4\pi} \operatorname{Im} W_{2f}(x, Q^2), \tag{18.146}$$

where W_{2f} is the second form factor of $W_f^{\mu\nu}$, defined in (18.144), neglecting terms suppressed by powers of Q^2. In the parton model evaluation,

$$f_f^+(x) = f_f(x) + f_{\bar{f}}(x). \tag{18.147}$$

From (18.123) and the definition (18.146), we know that $f_f^+(x)$ enters in the correct way into the formula for the deep inelastic scattering cross section. However, parton distribution functions have other important properties, including the normalization conditions (17.36) and (17.39) and the evolution with Q^2 discussed in Section 17.6. We must now see whether we can derive these properties from (18.146) using the operator product expansion.

The Dispersion Integral

The operator product analysis has given us explicit expressions for W_1 and W_2 as a series in inverse powers of Q^2. In the following discussion, we will concentrate on the analysis of W_2. We must work out the relation of its series expansion to the observable deep inelastic scattering cross section. As in the discussion of Section 18.4, the OPE analysis naturally takes place in an unphysical kinematic region. To make the operator product expansion, we

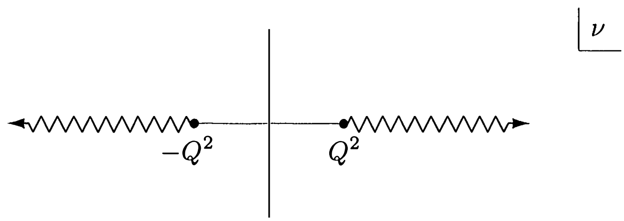


Figure 18.11. Analytic singularities of $W_2(\nu, Q^2)$ in the complex ν plane, for fixed Q^2.

needed to consider Q^2 to be larger than any other kinematic invariant. However, in the physical region for deep inelastic scattering, $2P \cdot q \geq Q^2$. We need a formula that connects these two distinct regions.

To state this problem more precisely, define

$$\nu = 2P \cdot q = ys; \tag{18.148}$$

in the frame in which the proton is at rest, $\nu = 2m_p q^0$. The form factor W_2 can be viewed as a function of ν and Q^2. Then, for fixed Q^2, the OPE gives a series expansion about the point $\nu = 0$, while the physical region for deep inelastic scattering is $\nu \geq Q^2$. Because this region is associated with a physical scattering process, $W_2(\nu, Q^2)$, viewed as an analytic function of ν for fixed Q^2, will have a branch cut along the real ν axis in this region. The discontinuity across this branch cut will be $(2i)$ times the imaginary part of W_2, which appears in the expression (18.123) for the deep inelastic cross section. Because expression (18.102) is symmetric under the interchange of (q, μ) and $(-q, \nu)$, W_2 must obey

$$W_2(-\nu, Q^2) = W_2(\nu, Q^2). \tag{18.149}$$

Thus, W_2 must also have a branch cut along the negative real axis, from $\nu = -Q^2$ to $-\infty$. The discontinuity across this cut gives the cross section for the u-channel process in which positive energy comes in through the second current and out through the first. Since $q^2 = -Q^2 < 0$, there is no possible physical t-channel process; thus W_2 has no further singularities in the complex ν plane. The analytic structure of $W_2(\nu, Q^2)$ is shown in Fig. 18.11.

Now consider the contour integral

$$I_n = \int \frac{d\nu}{2\pi i} \frac{1}{\nu^{n-1}} W_2(\nu, Q^2), \tag{18.150}$$

for n even, taken on a small circle surrounding the origin. This integral picks out the coefficient of ν^{n-2} in the series expansion for W_2. The OPE formula (18.145) gives us the leading contribution to this coefficient for large Q^2:

$$I_n = \sum_f Q_f^2 \frac{8}{(Q^2)^{n-1}} A_f^n. \tag{18.151}$$

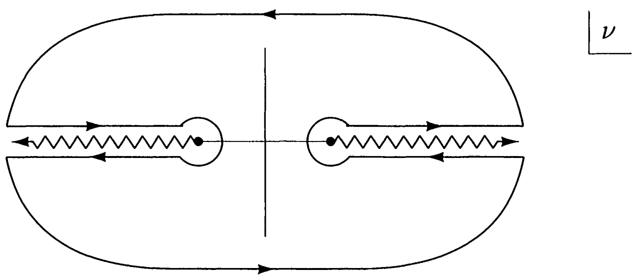


Figure 18.12. Contour of integration involved in the derivation of the moment sum rules for W_2.

The corrections to this formula are of order $\alpha_s(Q^2)$, from the evaluation of the OPE coefficient functions.

On the other hand, we can also distort the contour as shown in Fig. 18.12 and evaluate it as an integral over the discontinuities of W_2. By the symmetry (18.149), the two branch cuts give equal contributions. Thus,

$$I_n = 2 \int_{Q^2}^{\infty} \frac{d\nu}{2\pi i} \frac{1}{\nu^{n-1}} (2i) \operatorname{Im} W_2(\nu, Q^2). \tag{18.152}$$

Now change variables to $x = Q^2/\nu$. The integral becomes

$$I_n = \frac{8}{(Q^2)^{n-1}} \int_0^1 dx\, x^{n-2} \frac{\nu}{4\pi} \operatorname{Im} W_2. \tag{18.153}$$

When we equate (18.151) and (18.153) and relate $\operatorname{Im} W_2$ to the parton distributions $f_f^+(x)$ using (18.146), the relation we have derived splits into a series of sum rules,

$$\int_0^1 dx\, x^{n-1} f_f^+(x, Q^2) = A_f^n, \tag{18.154}$$

for n even. These relations are known as the *moment sum rules* for the deep inelastic form factors. They relate the x moments of the parton distribution functions, as defined by Eq. (18.146), to the proton matrix elements of twist-2 operators.

Because W_2 is a symmetric function of ν, the moment sum rules apply only for even n. However, in deep inelastic neutrino scattering, there is a third form factor in $W^{\mu\nu}$, associated with the interference term between the vector and axial vector parts of the weak interaction current. In Problem 18.2, we show that this form factor can be used to derive a set of sum rules for odd n:

$$\int_0^1 dx\, x^{n-1} f_f^-(x, Q^2) = A_f^n, \tag{18.155}$$

where A_f^n is the coefficient of the proton matrix element (18.134) for odd n, and $f_f^-(x)$ is a form factor which, in the parton model, evaluates to

$$f_f^-(x) = f_f(x) - f_{\bar{f}}(x). \tag{18.156}$$

Combining this information with the argument given below (18.136), we can see that the definition of the parton distribution functions from the deep inelastic form factors has the correct normalization. Using (18.137), we find

$$\int_0^1 dx\, f^-(x) = N_f, \tag{18.157}$$

the (net) number of quarks of flavor f in the proton. Similarly, (18.154) and (18.138) imply

$$\int_0^1 dx\, x f^+(x) = \langle x \rangle_f\,, \tag{18.158}$$

where $\langle x \rangle_f$ is the fraction of the total energy-momentum of the proton carried by quarks and antiquarks of flavor f.

Operator Rescaling

If the coefficients A_f^n were truly constants, relations (18.154) and (18.155) would be consistent with parton distribution functions that satisfy Bjorken scaling. However, as we remarked below (18.134), these factors actually depend on Q^2, since this is the normalization point of the operators in the operator product expansion (18.133). Since this dependence comes only through operator rescaling, it involves only logarithms of Q^2, and so contributes only a slow violation of Bjorken scaling. We can work out the Q^2 dependence of the parton distribution functions quantitatively by summing the leading logarithmic corrections to the matrix elements of the twist-2 operators.

To account for these corrections, let us first assume (incorrectly, as we will see below) that the twist-2 operators (18.132) are renormalized without operator mixing. Then the leading logarithmic corrections to the matrix element of the operator $\mathcal{O}_f^{(n)}$ would be summed by rescaling the operator normalized at Q to operators normalized at a standard reference point μ, of order 1 GeV. The relation between these conventions would be

$$\left[\mathcal{O}_f^{(n)}\right]_Q = \left(\frac{\log(Q^2/\Lambda^2)}{\log(\mu^2/\Lambda^2)}\right)^{a_f^n/2b_0} \left[\mathcal{O}_f^{(n)}\right]_\mu, \tag{18.159}$$

where a_f^n is the first coefficient of the γ function of $\mathcal{O}_f^{(n)}$. Then the factors A_f^n would depend on Q^2 according to

$$A_f^n(Q^2) = \left(\frac{\log(Q^2/\Lambda^2)}{\log(\mu^2/\Lambda^2)}\right)^{a_f^n/2b_0} A_f^n(\mu^2). \tag{18.160}$$

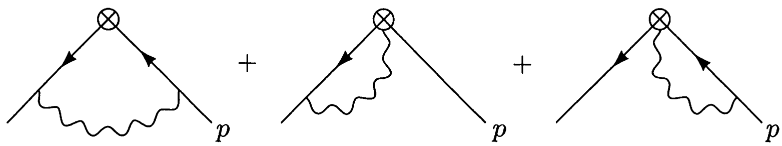
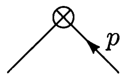


Figure 18.13. Diagrams contributing the anomalous dimension of the quark twist-2 operators.

This equation agrees with the scale dependence of operator product coefficients written in (18.76), for the special case of an operator product of currents, $a_1 = a_2 = 0$. To find the explicit form of the rescaling factor, we must compute a_f^n.

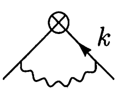
To compute the γ functions of the quark twist-2 operators, we must compute their counterterms for operator rescaling. These are determined by the diagrams shown in Fig. 18.13. It suffices to compute these diagrams with external momentum p entering through the quark line and zero external momentum injected into the operator. Under these conditions, the matrix element of the operator $\mathcal{O}_f^{(n)}$, in leading order, equals



$$= \gamma^{\mu_1} p^{\mu_2} \cdots p^{\mu_n}. \tag{18.161}$$

Here and at all later points in the discussion, we will treat the matrix elements of $\mathcal{O}_f^{(n)}$ as though they are symmetrized in the n indices and have all possible traces subtracted. We must now evaluate the diagrams of Fig. 18.13 and collect all terms that rescale this structure.

The first diagram of Fig. 18.13 is quite straightforward to evaluate:



$$\begin{aligned} &= \int \frac{d^4k}{(2\pi)^4} (ig)^2 \gamma^\lambda t^a \frac{i\not{k}}{k^2} \gamma^{\mu_1} k^{\mu_2} \cdots k^{\mu_n} \frac{i\not{k}}{k^2} \gamma_\lambda t^a \frac{-i}{(k-p)^2} \\ &= -ig^2 C_2(r) \cdot (-2) \int \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2)^2 (k-p)^2} \not{k} \gamma^{\mu_1} k^{\mu_2} \cdots k^{\mu_n} \not{k}. \end{aligned} \tag{18.162}$$

We combine denominators using identity (6.40):

$$\frac{1}{(k^2)^2 (k-p)^2} = \int_0^1 dx \, \frac{2(1-x)}{(\mathbf{k}^2 - \Delta)^3}; \tag{18.163}$$

the quantities in the denominator on the right are $\mathbf{k} = k - xp$ and $\Delta = -x(1-x)p^2$. We must now shift the integral, substitute $k = \mathbf{k} + xp$ in the numerator, and pick out a term proportional to $(n-1)$ powers of p. If this term contains the factor $g^{\mu_i \mu_j}$, we may drop it, since it contributes to the coefficient of an operator of higher twist and since, in any event, it will be removed when we subtract traces. Thus, we must choose carefully which two factors of k we replace with $\mathbf{k}$ when we replace the others with (xp). The

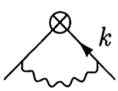
following choices, simplified using the rotational symmetry of the **k** integral, do not give useful contributions:

$$\begin{aligned} \mathbf{k}^{\mu_i}\mathbf{k}^{\mu_j} &= \frac{1}{4}\mathbf{k}^2 g^{\mu_i\mu_j}, \\ \not{\mathbf{k}}\gamma^{\mu_1}\mathbf{k}^{\mu_j} &\to \frac{1}{4}\mathbf{k}^2\gamma^{\mu_j}\gamma^{\mu_1} = \frac{1}{4}\mathbf{k}^2 g^{\mu_1\mu_j}. \end{aligned} \tag{18.164}$$

In the second line, we have used the symmetry under $\mu_1 \leftrightarrow \mu_j$. The one remaining placement of the factors of **k** is

$$\not{\mathbf{k}}\gamma^{\mu_1}\not{\mathbf{k}} \to \frac{1}{4}\mathbf{k}^2\gamma^{\nu}\gamma^{\mu_1}\gamma_{\nu} = -\frac{1}{2}k^2\gamma^{\mu_1}. \tag{18.165}$$

Thus (18.162) has the value



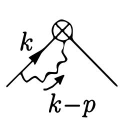
$$\begin{aligned} &= -ig^2\frac{4}{3}\int_0^1 dx\cdot 2(1-x)\int\frac{d^4\mathbf{k}}{(2\pi)^4}\frac{\mathbf{k}^2}{(\mathbf{k}^2-\Delta)^3}\gamma^{\mu_1}(xp^{\mu_2})\cdots(xp^{\mu_n}) \\ &= -i\frac{8}{3}g^2\int_0^1 dx(1-x)x^{n-1}\frac{i}{(4\pi)^2}\Gamma(2-\tfrac{d}{2})\gamma^{\mu_1}p^{\mu_2}\cdots p^{\mu_n} \\ &= \frac{4}{3}\frac{2}{n(n+1)}\frac{g^2}{(4\pi)^2}\Gamma(2-\tfrac{d}{2})\gamma^{\mu_1}p^{\mu_2}\cdots p^{\mu_n}. \end{aligned} \tag{18.166}$$

It is not so obvious that there are additional contributions to the rescaling of the operators $\mathcal{O}_f^{(n)}$. Note, however, that the covariant derivatives in (18.132) contain explicit factors of the gauge field,

$$iD^{\mu_j} = i\partial^{\mu_j} - gA^{a\mu_j}t^a, \tag{18.167}$$

and these may be contracted with gauge field vertices on the external legs. These contributions give rise to the second and third diagrams in Figure 18.13. The term in which two factors of $A^{a\mu}$ from (18.167) are contracted with one another is proportional to $G^{\mu_i\mu_j}$ and thus does not contribute to the rescaling of the leading-twist operators.

The contributions we have just described have the form of sums over j, where μ_j is the index of the derivative that includes the contraction. Then the second diagram of Fig. 18.3 is the sum over j of the following integral:

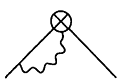


$$\begin{aligned} &= \int\frac{d^4k}{(2\pi)^4}(ig)\gamma_\lambda t^a\frac{i\not{k}}{k^2}\gamma^{\mu_1}k^{\mu_2}\cdots k^{\mu_{j-1}} \\ &\qquad\cdot(-gt^a g^{\lambda\mu_j})p^{\mu_{j+1}}\cdots p^{\mu_n}\frac{-i}{(k-p)^2} \\ &= ig^2C_2(r)\int\frac{d^4k}{(2\pi)^4}\frac{1}{k^2(k-p)^2}\gamma^{\mu_j}\not{k}\gamma^{\mu_1}k^{\mu_2}\cdots k^{\mu_{j-1}}p^{\mu_{j+1}}\cdots p^{\mu_n}. \end{aligned} \tag{18.168}$$

Since μ_j and μ_1 are symmetrized, we can use (18.128) to rewrite

$$\begin{aligned}\gamma^{\mu_j} \not{k} \gamma^{\mu_1} &\rightarrow k^{\mu_j}\gamma^{\mu_1} + \gamma^{\mu_j} k^{\mu_1} - g^{\mu_j \mu_1} \not{k} \\ &\rightarrow 2\gamma^{\mu_1} k^{\mu_j},\end{aligned} \tag{18.169}$$

where, in the second line, the symmetrization of indices and subtraction of traces is understood. Now combine denominators. To obtain a term with $(n-1)$ factors of p, we must replace every factor k in the numerator of (18.168) with (xp). This gives



$$\begin{aligned} &= ig^2 \frac{4}{3} \int_0^1 dx \int \frac{d^4\mathbf{k}}{(2\pi)^4} \frac{1}{(\mathbf{k}^2 - \Delta)^2} 2\gamma^{\mu_1}(xp^{\mu_2}) \cdots (xp^{\mu_j}) p^{\mu_{j+1}} \cdots p^{\mu_n} \\ &= ig^2 \frac{8}{3} \int_0^1 dx\, x^{j-1} \frac{i}{(4\pi)^2} \Gamma(2-\tfrac{d}{2}) \gamma^{\mu_1} p^{\mu_2} \cdots p^{\mu_n} \\ &= -\frac{4}{3}\frac{2}{j}\frac{g^2}{(4\pi)^2} \Gamma(2-\tfrac{d}{2}) \gamma^{\mu_1} p^{\mu_2} \cdots p^{\mu_n}. \end{aligned} \tag{18.170}$$

This contribution must be summed over j from 2 to n. The third diagram of Fig. 18.13 makes an equal contribution.

Summing the rescaling factors from the three diagrams of Fig. 18.13, we find for the operator rescaling counterterm of $\mathcal{O}_f^{(n)}$

$$\delta_f = \frac{g^2}{(4\pi)^2}\frac{4}{3}\left[4\sum_{j=2}^{n}\frac{1}{j} - \frac{2}{n(n+1)}\right]\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \tag{18.171}$$

From this result, we can derive the Callan-Symanzik γ function by the use of (18.23) and the field strength renormalization counterterm (18.9). We find

$$\gamma_f^n = \frac{8}{3}\frac{g^2}{(4\pi)^2}\left[1 + 4\sum_{2}^{n}\frac{1}{j} - \frac{2}{n(n+1)}\right]. \tag{18.172}$$

Notice that this expression vanishes for $n = 1$, so that there is no rescaling of A_f^1, as required by (18.157). For $n > 1$, γ_f^n is positive and so its coefficient a_f^n is negative. This implies that the higher moments of the quark distribution functions are suppressed as Q^2 becomes large.

Operator Mixing

The QCD rescaling of the operators $\mathcal{O}_f^{(n)}$ is still more complicated because QCD contains additional twist-2 operators which can be built from gluon fields. These new operators are mixed with the quark twist-2 operators by the diagrams of Fig. 18.14.

For n even, the diagrams of Fig. 18.14 give the operators $\mathcal{O}_f^{(n)}$ matrix elements in the state of a gluon with momentum p. The tensor structure of

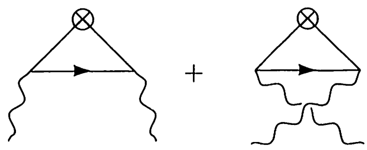


Figure 18.14. Diagrams that produce operator mixing between twist-2 quark and gluon operators.

this matrix element contains the term

$$g^{\alpha\beta}p^{\mu_1}\cdots p^{\mu_n}, \tag{18.173}$$

where α, β are the polarization indices of the external gluons. This structure arises from the operator

$$\mathcal{O}_g^{(n)\mu_1\cdots\mu_n} = -F^{\{\mu_1\nu}(iD^{\mu_2})\cdots(iD^{\mu_{n-1}})F^{\mu_n\}}{}_\nu - \text{traces}, \tag{18.174}$$

symmetrized on $\mu_1, \cdots \mu_n$, with traces subtracted. These operators have dimension $(n+2)$ and spin n, and thus have twist 2.

The gluon operators (18.174) are relevant only for n even. Using the manipulation

$$F^{\mu_1\nu}(iD^{\mu_2})\cdots F^{\mu_n}{}_\nu = i\partial^{\mu_2}\big(F^{\mu_1\nu}\cdots F^{\mu_n}{}_\nu\big) - (iD^{\mu_2})F^{\mu_1\nu}\cdots F^{\mu_n}{}_\nu, \tag{18.175}$$

we can transfer the covariant derivatives from one factor of $F^{\mu\nu}$ to the other, giving

$$\mathcal{O}_g^{(n)} = (-1)^n\mathcal{O}_g^{(n)} + \partial^{\mu_1}(\mathcal{O}'). \tag{18.176}$$

Thus, for n odd, the operator $\mathcal{O}_g^{(n)}$ is equal to a total derivative. The matrix elements of a total derivative are proportional to the momentum injected into this operator. Since zero momentum is injected in the calculation of the proton matrix elements of the OPE of currents, the operators $\mathcal{O}_g^{(n)}$ have no effect on the deep inelastic scattering cross section for n odd.

For n even, however, we must take account of the mixing of $\mathcal{O}_g^{(n)}$ with $\mathcal{O}_f^{(n)}$. The computation of the diagrams of Fig. 18.14 is quite similar to the other operator rescaling calculations we have done in this chapter, and so we reserve working out the details for Problem 18.3. We find that the diagrams of Fig. 18.14 contain a structure proportional to (18.173) with the coefficient

$$\frac{2(n^2+n+2)}{n(n+1)(n+2)}\frac{g^2}{(4\pi)^2}\Gamma(2-\tfrac{d}{2}). \tag{18.177}$$

From this computation, we find that the renormalized twist-2 quark operator, properly normalized at the scale M, is given in terms of bare operators by

$$\big[\mathcal{O}_f^{(n)}\big]_M = (1+\delta_f)\big[\mathcal{O}_f^{(n)}\big]_0 + (\delta_g)\big[\mathcal{O}_g^{(n)}\big]_0, \tag{18.178}$$

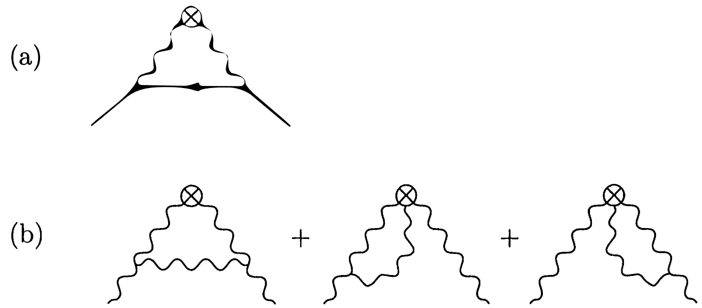


Figure 18.15. Diagrams contributing to the operator rescaling of twist-2 gluon operators: (a) contributions to gluon-quark mixing; (b) contributions to diagonal gluon operator renormalization.

where δ_f is given by (18.171) and

$$\delta_g = -\frac{g^2}{(4\pi)^2}\frac{2(n^2+n+2)}{n(n+1)(n+2)}\frac{\Gamma(2-\frac{d}{2})}{(M^2)^{2-d/2}}. \tag{18.179}$$

This equation gives us two elements of the anomalous dimension matrix of twist-2 operators.

The remaining elements of the γ matrix for twist-2 operators are generated by the diagrams shown in Fig. 18.15. The diagram of Fig. 18.15(a) gives the mixing of $\mathcal{O}_g^{(n)}$ back into $\mathcal{O}_f^{(n)}$. The diagrams of Fig. 18.15(b), combined with the counterterm δ_3 for gluon field strength rescaling, gives the diagonal anomalous dimension. The counterterm δ_3 is given explicitly, in Feynman-'t Hooft gauge, in (16.74). The remainder of this anomalous dimension computation is discussed in Problem 18.3.

To describe the complete anomalous dimension matrix, we begin by considering a strong interaction model with one quark flavor. In this case, there is one twist-two operator $\mathcal{O}_f^{(n)}$ which mixes with $\mathcal{O}_g^{(n)}$. These two operators mix through a 2×2 matrix

$$\gamma^n = -\frac{g^2}{(4\pi)^2}\begin{pmatrix} a^n_{ff} & a^n_{fg} \\ a^n_{gf} & a^n_{gg}\end{pmatrix}, \tag{18.180}$$

where

$$\begin{aligned}
a^n_{ff} &= -\frac{8}{3}\Big[1+4\sum_2^n\frac{1}{j}-\frac{2}{n(n+1)}\Big],\\
a^n_{fg} &= 4\frac{n^2+n+2}{n(n+1)(n+2)},\\
a^n_{gf} &= \frac{16}{3}\frac{n^2+n+2}{n(n^2-1)},\\
a^n_{gg} &= -6\Big[\frac{1}{3}+\frac{2}{9}n_f+4\sum_2^n\frac{1}{j}-\frac{4}{n(n-1)}-\frac{4}{(n+1)(n+2)}\Big].
\end{aligned} \tag{18.181}$$

Notice that this matrix is not symmetric. In the last line, n_f is the number of quark flavors, equal to 1 in this case; this term comes from (16.74).

In the realistic case, QCD contains several quark flavors—u, d, s, and also c and b when we work at momenta sufficiently large that we can ignore the masses of these particles. Then the anomalous dimension matrix γ^n has size $(n_f+1)\times(n_f+1)$. The submatrix acting on quark operators is diagonal, with all of the diagonal entries being given by a^n_{ff} in (18.181). The quark-gluon and gluon-quark entries are all given by a^n_{fg} and a^n_{gf}, respectively, and are independent of the flavor. The gluon diagonal entry is given by a^n_{gg} in (18.181) with the realistic value of n_f. This means that the gluon operator mixes with only one linear combination of quark operators:

$$\sum_f \mathcal{O}_f^{(n)}; \tag{18.182}$$

the orthogonal linear combinations are simply rescaled, with the exponent given by a^n_{ff} or (18.172).

Let us now apply this analysis of operator mixing to the evaluation of the moment sum rules. For odd n, there is no operator mixing, and so the Q^2 dependence of the right-hand side of (18.155) is correctly given by the simple rescaling (18.160).

For even n, we must take operator mixing into account. The right-hand side of the sum rule (18.154) is the proton matrix element of a twist-2 operator normalized at the scale Q. Let us write an arbitrary linear combination of these operators as

$$c_i^n\big[\mathcal{O}_i^{(n)}\big]_Q, \tag{18.183}$$

where the index i runs over g and the various flavors f. To rescale this operator to a fixed reference momentum μ, we rewrite the coefficients in a basis of *left* eigenvectors of γ^n and rescale each eigenvector acccording to (18.159). In terms of the matrix a^n_{ij} of rescaling coefficients, we can write the rescaling abstractly as

$$c_i^n\big[\mathcal{O}_i^{(n)}\big]_Q = c_i^n\left\{\left(\frac{\log(Q^2/\Lambda^2)}{\log(\mu^2/\Lambda^2)}\right)^{a^n/2b_0}\right\}_{ij}\big[\mathcal{O}_j^{(n)}\big]_\mu. \tag{18.184}$$

This rescaling, acting with c_i^n to the left of the matrix (a^n), is precisely the prescription required by Eq. (18.79).

Let us work this out explicitly for the case $n=2$. The right-hand side of the moment sum rule (18.154) is given by the matrix element of $\mathcal{O}_f^{(2)}$. We rewrite this as

$$\mathcal{O}_f^{(2)} = \big(\mathcal{O}_f^{(2)} - \frac{1}{n_f}\sum_{f'}\mathcal{O}_{f'}^{(2)}\big) + \frac{1}{n_f}\sum_{f'}\mathcal{O}_{f'}^{(2)}. \tag{18.185}$$

The first term is simply rescaled; the second term mixes with the gluon operator $\mathcal{O}_g^{(2)}$. The anomalous dimension matrix acting on $(\sum_f \mathcal{O}_f, \mathcal{O}_g)$ for $n=2$

has coefficients

$$\begin{pmatrix} a^2_{ff} & a^2_{fg} n_f \\ a^2_{gf} & a^2_{gg} \end{pmatrix} = \begin{pmatrix} -\frac{64}{9} & \frac{4}{3} n_f \\ \frac{64}{9} & -\frac{4}{3} n_f \end{pmatrix}. \tag{18.186}$$

The left eigenvectors of this matrix, and their corresponding eigenvalues, are

$$\begin{aligned} (1,1) \quad &\to \quad a^2 = 0 \\ \left(\frac{16}{3}, -n_f\right) \quad &\to \quad a^2 = -\frac{4}{3}\left(\frac{16}{3} + n_f\right). \end{aligned} \tag{18.187}$$

Notice that the first eigenvector gives a linear combination of operators $c_i^2 \mathcal{O}_i^{(2)}$ with zero anomalous dimension. This operator is in fact the total energy momentum tensor of QCD,

$$T^{\mu\nu} = \sum_f \mathcal{O}_f^{(2)\,\mu\nu} + \mathcal{O}_g^{(2)\,\mu\nu}, \tag{18.188}$$

which must have $\gamma = 0$. If we expand the second term in (18.185) in terms of the components (18.187), we can compute the full form of the operator rescaling. We find

$$\begin{aligned} \left[\mathcal{O}_f^{(2)}\right]_Q = &\frac{1}{16/3 + n_f} T \\ &+ \frac{1}{n_f(\frac{16}{3} + n_f)} \left(\frac{\log(Q^2/\Lambda^2)}{\log(\mu^2/\Lambda^2)}\right)^{-\frac{4}{3}(\frac{16}{3}+n_f)/2b_0} \left[\frac{16}{3} \sum_f \mathcal{O}_f^{(2)} - n_f \mathcal{O}_g^{(2)}\right]_\mu \\ &+ \left(\frac{\log(Q^2/\Lambda^2)}{\log(\mu^2/\Lambda^2)}\right)^{-32/9b_0} \left[\mathcal{O}_f^{(2)} - \frac{1}{n_f} \sum_{f'} \mathcal{O}_{f'}^{(2)}\right]_\mu, \end{aligned} \tag{18.189}$$

where T is the energy-momentum tensor (18.188). The right-hand side of the $n = 2$ moment sum rule is given by the coefficient of the proton matrix element of this operator. To evaluate this coefficient, we need to define gluon analogues of the A_f^n, by writing, analogously to (18.134),

$$\langle P | \, \mathcal{O}_g^{(n)\mu_1 \cdots \mu_n} \, | P \rangle = A_g^n \cdot 2P^{\mu_1} \cdots P^{\mu_n} - \text{traces}. \tag{18.190}$$

For the case $n = 2$, we note in particular that

$$\langle P | \, T^{\mu\nu} \, | P \rangle = 2P^\mu P^\nu; \tag{18.191}$$

thus, (18.188) implies

$$\sum_f A_f^2 + A_g^2 = 1. \tag{18.192}$$

If we replace each operator in (18.189) by the corresponding coefficient $A_i^2(\mu)$, we will have an expression for the right-hand side of the $n = 2$ moment sum rule which makes its Q^2 dependence explicit.

Although expression (18.189) is rather complicated, it has a simple form in the extreme limit $Q^2 \to \infty$. At asymptotic Q^2, the last two terms of (18.189)

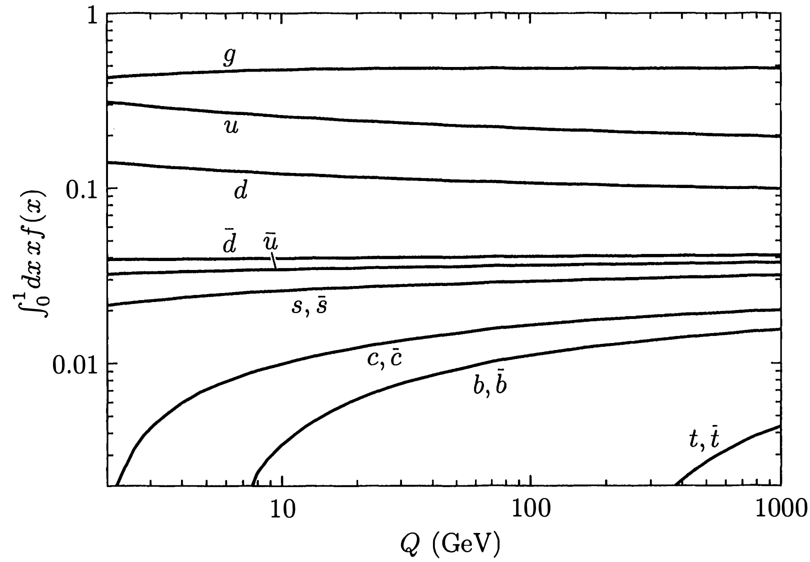


Figure 18.16. Fractions of the total energy-momentum of the proton carried by various parton species, as a function of Q, according to the CTEQ fit to deep inelastic scattering data described in Fig. 17.6. The Q dependence of the curves is calculated from the QCD evolution equations.

tend to zero, and the right-hand side of (18.189) becomes a fixed number times the energy-momentum tensor. Then, using (18.191), we can evaluate the $n = 2$ moment sum rule completely:

$$\int_0^1 dx\, x f_f^+(x) \to \frac{1}{16/3 + n_f}. \tag{18.193}$$

In this extreme limit, we find that each quark flavor carries the same fixed fraction of the energy-momentum of the proton. By (18.192), the remainder is carried by the gluons. To illustrate, in a theory with $n_f = 4$, each quark flavor carries $3/28$ of the total momentum of the proton, and the gluons carry the remaining $4/7$. Figure 18.16 shows how slowly these asymptotic results are approached starting from realistic parton distributions.

Relation to the Altarelli-Parisi Equations

The operator mixing analysis just described gives predictions for the moments of parton distributions which imply that these integrals are Q^2 dependent. Of the various moment integrals that do not involve operator mixing, only the $n = 1$ integrals which give the flavor quantum numbers of the proton are constant as a function of Q^2. The rest decrease as powers of $\log Q^2$. Similarly, one linear combination of the matrix elements of $n = 2$ twist-2 operators

remains constant with Q^2. This relation is expression by the sum rule (18.192). To write this relation more clearly, let us introduce the parton distribution of gluons as a smooth function satisfying the relations

$$\int_0^1 dx\, x^{n-1} f_g(x, Q^2) = A_g^{(n)}(Q^2). \tag{18.194}$$

Then (18.192) becomes just the total momentum sum rule for parton distributions (17.39):

$$\int_0^1 dx\, x \Big[\sum_f f_f^+(x) + f_g(x)\Big] = 1. \tag{18.195}$$

It is not difficult to verify that, for $n > 2$, all of the eigenvalues of the matrix a_{ij}^n of anomalous dimension coefficients are negative. Thus, all of the higher moment sum rules decrease, subject to the flavor charge and momentum conservation laws. In other words, the operator renormalization analysis predicts that parton distributions shift down to smaller values of x as $\log Q^2$ increases. It is pleasing that this is the same conclusion that we reached in Section 17.5, where we derived the Altarelli-Parisi equations to describe this evolution of the parton distributions.

Given that the operator analysis and the Altarelli-Parisi equations imply the same qualitative behavior for the parton distributions, how do these analyses compare quantitatively? To compare them directly, we should work out what predictions the Altarelli-Parisi equations make for the moments of the parton distribution functions. Let us begin with the simpler case of $f_f^-(x) = f_f(x) - f_{\bar{f}}(x)$.

To find the Altarelli-Parisi equation for this quantity, subtract the last two equations of (17.128). The term involving the gluon distribution cancels, and we find

$$\frac{d}{d\log Q^2} f_f^-(x) = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{q\leftarrow q}(z) f_f^-\big(\frac{x}{z}\big). \tag{18.196}$$

Now define

$$M_{fn}^- = \int_0^1 dx\, x^{n-1} f_f^-(x). \tag{18.197}$$

This quantity obeys the differential equation

$$\frac{d}{d\log Q^2} M_{fn}^- = \frac{\alpha_s(Q^2)}{2\pi} \int_0^1 dx\, x^{n-1} \int_x^1 \frac{dz}{z} P_{q\leftarrow q}(z) f_f^-\big(\frac{x}{z}\big). \tag{18.198}$$

Interchange the order of integration on the right-hand side, and change variables to $y = x/z$:

$$\begin{aligned}\int_0^1 dx\, x^{n-1} \int_x^1 \frac{dz}{z} &= \int_0^1 \frac{dz}{z} \int_0^z dx\, x^{n-1}\\ &= \int_0^1 \frac{dz}{z} \int_0^1 dy\, y^{n-1} z^n\\ &= \int_0^1 dz\, z^{n-1} \int_0^1 dy\, y^{n-1}.\end{aligned} \tag{18.199}$$

Then the right-hand side of the differential equation neatly factorizes:

$$\frac{d}{d\log Q^2} M_{nf}^- = \frac{\alpha_s(Q^2)}{2\pi}\left[\int_0^1 dz\, z^{n-1} P_{q\leftarrow q}(z)\right] \cdot \int_0^1 dy\, y^{n-1} f_f^-(y); \tag{18.200}$$

the last factor is again M_{fn}^-. The coefficient in this relation is the nth moment of the splitting function $P_{q\leftarrow q}(z)$. We can compute this from the explicit form of this function given in (17.129):

$$\int_0^1 dz\, z^{n-1} P_{q\leftarrow q}(z) = \int_0^1 dz\, z^{n-1} \frac{4}{3}\left[\frac{1+z^2}{(1-z)_+} + \frac{3}{2}\delta(1-z)\right]. \tag{18.201}$$

The integral over the distribution is done by using the definition (17.105):

$$\begin{aligned}\int_0^1 dz\, z^{n-1} \frac{1}{(1-z)_+} &= \int_0^1 dz\, \frac{z^{n-1}-1}{(1-z)}\\ &= \int_0^1 dz(-1 - z - \cdots - z^{n-2})\\ &= -\sum_1^{n-1} \frac{1}{j}.\end{aligned} \tag{18.202}$$

Then

$$\begin{aligned}\int_0^1 dz\, z^{n-1} P_{q\leftarrow q}(z) &= -\frac{4}{3}\left[\sum_1^{n-1}\frac{1}{j} + \sum_1^{n+1}\frac{1}{j} - \frac{3}{2}\right]\\ &= -\frac{2}{3}\left[1 + 4\sum_2^{n}\frac{1}{j} - \frac{2}{n(n+1)}\right].\end{aligned} \tag{18.203}$$

Remarkably, this is just $a_f^n/4$, as the anomalous dimension coefficient is given in (18.172) or (18.181). Thus, according to the Altarelli-Parisi equations, the nth moment of $f_f^-(x)$ obeys

$$\frac{d}{d\log Q^2}M_{fn}^- = \frac{\alpha_s(Q^2)}{8\pi}a_f^n \cdot M_{fn}^-. \tag{18.204}$$

To integrate this equation, we need the explicit form of $\alpha_s(Q^2)$. Inserting expression (17.17), we find

$$\frac{d}{d\log Q^2}M_{fn}^- = \frac{a_f^n}{2b_0}\frac{1}{\log(Q^2/\Lambda^2)}M_{fn}^-. \tag{18.205}$$

The solution of this equation, derived from the Altarelli-Parisi equations, is precisely the function (18.160) that we derived from the operator analysis of the moment sum rules for f_f^-.

It is not difficult to check that this conclusion is more general. By taking the nth moment of the full Altarelli-Parisi equations (17.128), we convert these equations to a set of ordinary differential equations for the moments. The linear combination of quark distribution functions

$$\sum_f \left(f_f(x) + f_{\bar f}(x)\right) \tag{18.206}$$

couples to the gluon distribution and leads to a 2×2 set of equations. All orthogonal linear combinations separate from the gluon distribution and thus have moments that obey equations identical to (18.205). To analyze the coupled equations, define

$$M_n^+ = \int_0^1 dx\, x^{n-1}\sum_f \left(f_f(x) + f_{\bar f}(x)\right); \qquad M_{gn} = \int_0^1 dx\, x^{n-1} f_g(x). \tag{18.207}$$

Then one can show, by the manipulations that led to (18.205), that the Altarelli-Parisi equations predict for these moments the set of coupled equations

$$\begin{aligned}\frac{d}{d\log Q^2}M_n^+ &= \frac{1}{2b_0}\frac{1}{\log(Q^2/\Lambda^2)}\left[a_{ff}^n M_n^+ + a_{fg}^n M_{gn}\right],\\ \frac{d}{d\log Q^2}M_{gn} &= \frac{1}{2b_0}\frac{1}{\log(Q^2/\Lambda^2)}\left[n_f a_{gf}^n M_n^+ + a_{gg}^n M_{gn}\right],\end{aligned} \tag{18.208}$$

where the coefficients a_{ij}^n are proportional to the nth moments of the splitting functions given in (17.129) and (17.130). In all cases, one can see that these coefficients agree precisely with the corresponding coefficients in (18.181). Thus, the solution of these equations gives the same Q^2 dependence for the moments of parton distribution functions that we found from the operator analysis.

Remarkably, the analysis of parton splitting functions given in Chapter 17 and the analysis of operator renormalization factors given above have turned out to be two views of the same basic phenomenon. Both sets of equations

express the manner in which the constituents of hadrons in QCD are resolved, layer by layer, by hard-scattering processes at successively higher values of the momentum transfer. Our understanding that a quark, when studied on a fine scale, is resolved into a set of quarks, antiquarks, and gluons indicates that we have gone far beyond the simple notions of one-particle relativistic mechanics. Our two complementary derivations of this idea reinforce its fundamental character as a prediction of quantum field theory. It is especially pleasing that, as we saw at the end of Section 17.5, Nature apparently accepts this prediction and makes this consequence of quantum field theory an essential part of the structure of hadrons.

Problems

18.1 Matrix element for proton decay. Some advanced theories of particle interactions include heavy particles X whose couplings violate the conservation of baryon number. Integrating out these particles produces an effective interaction that allows the proton to decay to a positron and a photon or a pion. This effective interaction is most easily written using the definite-helicity components of the quark and electron fields: If u_L, d_L, u_R, e_R are two-component spinors, then this effective interaction is

$$\Delta\mathcal{L} = \frac{2}{m_X^2}\epsilon_{abc}\epsilon^{\alpha\beta}\epsilon^{\gamma\delta}e_{R\alpha}u_{Ra\beta}u_{Lb\gamma}d_{Lc\delta}.$$

A typical value for the mass of the X boson is $m_X = 10^{16}$ GeV.

(a) Estimate, in order of magnitude, the value of the proton lifetime if the proton is allowed to decay through this interaction.

(b) Show that the three-quark operator in $\Delta\mathcal{L}$ has an anomalous dimension

$$\gamma = -4\frac{g^2}{(4\pi)^2}.$$

Estimate the enhancement of the proton decay rate due to the leading QCD corrections.

18.2 Parity-violating deep inelastic form factor. In this problem, we first motivate the presence of additional deep inelastic form factors that are proportional to *differences* of quark and antiquark distribution functions. Then we define these functions formally and work out their properties.

(a) Analyze neutrino-proton scattering following the method used at the beginning of Section 18.5. Define

$$J_+^\mu = \bar{u}\gamma^\mu\left(\frac{1-\gamma^5}{2}\right)d, \qquad J_-^\mu = \bar{d}\gamma^\mu\left(\frac{1-\gamma^5}{2}\right)u.$$

Let

$$W^{\mu\nu(\nu)} = 2i\int d^4x\, e^{iq\cdot x}\left\langle P\right| T\{J_-^\mu(x)J_+^\nu(0)\}\left|P\right\rangle,$$

averaged over the proton spin. Show that the cross section for deep inelastic neutrino scattering can be computed from $W^{\mu\nu(\nu)}$ according to

$$\frac{d^2\sigma}{dxdy}(\nu p \to \mu^- X) = \frac{G_F^2 y}{2\pi^2} \cdot \mathrm{Im}\big[(k_\mu k'_\nu + k_\nu k'_\mu - g_{\mu\nu} k\cdot k' - i\epsilon_{\mu\nu}{}^{\alpha\beta} k'_\alpha k_\beta) W^{\mu\nu(\nu)}(P,q)\big].$$

(b) Show that any term in $W^{\mu\nu(\nu)}$ proportional to q^μ or q^ν gives zero when contracted with the lepton momentum tensor in the formula above. Thus we can expand $W^{\mu\nu(\nu)}$ with three scalar form factors,

$$W^{\mu\nu(\nu)} = -g^{\mu\nu} W_1^{(\nu)} + P^\mu P^\nu W_2^{(\nu)} + i\epsilon^{\mu\nu\lambda\sigma} P_\lambda q_\sigma W_3^{(\nu)} + \cdots,$$

where the additional terms do not contribute to the deep inelastic cross section. Find the formula for the deep inelastic cross section in terms of the imaginary parts of $W_1^{(\nu)}$, $W_2^{(\nu)}$, and $W_3^{(\nu)}$.

(c) Evaluate the form factors $W_i^{(\nu)}$ in the parton model, and show that

$$\begin{aligned}\mathrm{Im}\, W_1^{(\nu)} &= \pi(f_d(x) + f_{\bar{u}}(x)),\\ \mathrm{Im}\, W_2^{(\nu)} &= \frac{4\pi}{ys} x(f_d(x) + f_{\bar{u}}(x)),\\ \mathrm{Im}\, W_3^{(\nu)} &= \frac{2\pi}{ys}(f_d(x) - f_{\bar{u}}(x)).\end{aligned}$$

Insert these expressions into the formula derived in part (b) and show that the result reproduces the first line of Eq. (17.35).

(d) This analysis motivates the following definition: For a single quark flavor f, let

$$J^\mu_{fL} = \bar{f}\gamma^\mu\left(\frac{1-\gamma^5}{2}\right)f.$$

Define

$$W^{\mu\nu}_{fL} = 2i\int d^4x\, e^{iq\cdot x}\,\langle P|\,T\{J^\mu_{fL}(x)J^\nu_{fL}(0)\}\,|P\rangle\,.$$

Decompose this tensor according to

$$W^{\mu\nu}_{fL} = -g^{\mu\nu}W_{1fL} + P^\mu P^\nu W_{2fL} + i\epsilon^{\mu\nu\lambda\sigma}P_\lambda q_\sigma W_{3fL} + \cdots,$$

where the remaining terms are proportional to q^μ or q^ν. Evaluate the W_{iL} in the parton model. Show that the quantities W_{1fL} and W_{2fL} reproduce the expressions for W_{1f} and W_{2f} given by Eqs. (18.120) and (18.144), and that W_{3fL} is given by

$$\mathrm{Im}\, W_{3fL} = \frac{2\pi}{ys}(f_f(x) - f_{\bar{f}}(x)).$$

(e) Compute the operator product of the currents in the expression for $W^{\mu\nu}_{fL}$, and write the terms in this product that involve twist-2 operators. Show that the expressions for W_{1fL} and W_{2fL} that follow from this analysis reproduce the expressions for W_{1f} and W_{2f} given by Eqs. (18.144) and (18.145). Find the corresponding expression for W_{3fL}.

(f) Define the parton distribution f_f^- by the relation

$$f_f^-(x, Q^2) = \frac{ys}{2\pi} \operatorname{Im} W_{3fL}(x, Q^2).$$

Show that, by virtue of this definition, the distribution function f_f^- satisfies the sum rule (18.155) for odd n.

18.3 Anomalous dimensions of gluon twist-2 operators.

(a) Compute the divergent parts of the diagrams in Fig. 18.14, and use these to derive the second line of Eq. (18.181). Notice that this result holds only for n even. Show that the two diagrams cancel for n odd.

(b) Compute the divergent parts of the diagrams in Fig. 18.5, and use these to derive the third and fourth lines of Eq. (18.181).

18.4 Deep inelastic scattering from a photon. Consider the problem of deep-inelastic scattering of an electron from a photon. This process can actually be measured by analyzing the reaction $e^+e^- \to e^+e^- + X$ in the regime where the positron goes forward, with emission of a collinear photon, which then has a hard reaction with the electron. Let us analyze this process to leading order in QED and to leading-log order in QCD. To predict the photon structure functions, it is reasonable to integrate the renormalization group equations with the initial condition that the parton distribution for photons in the photon is $\delta(x-1)$ at $Q^2 = (\frac{1}{2}\text{ GeV})^2$. Take $\Lambda = 150$ MeV. Assume for simplicity that there are four flavors of quarks, u, d, c, and s, with charges 2/3, −1/3, 2/3, −1/3, respectively, and that it is always possible to ignore the masses of these quarks.

(a) Use the Altarelli-Parisi equations to compute the parton distributions for quarks and antiquarks in the photon, to leading order in QED and to zeroth order in QCD. Compute also the probability that the photon remains a photon as a function of Q^2.

(b) Formulate the problem of computing the moments of W_2 for the photon as a problem in operator mixing. Compute the relevant anomalous dimension matrix γ. You should be able to assemble this matrix from familiar ingredients without doing further Feynman diagram computations.

(c) Compute the $n = 2$ moments of the photon structure functions as a function of Q^2.

(d) Describe qualitatively the evolution of the photon structure function as a function of x and Q^2.

Chapter 19

Perturbation Theory Anomalies

In many examples, we have seen that loop corrections can have an important effect on the predictions of quantum field theory. We have studied examples in which the relative importance of operators is shifted by radiative corrections, and in which the form of the interactions they mediate is altered. However, in specific circumstances, radiative corrections can have an even more significant effect: They can destroy symmetries of the classical equations of motion.

The most important effect of this type involves the chiral symmetries of theories with massless fermions. In Section 3.4, we saw that the massless Dirac Lagrangian has an enhanced symmetry associated with the separate number conservation of left- and right-handed fermions. This symmetry is generated by the axial vector current $j^{\mu 5} = \bar{\psi}\gamma^\mu\gamma^5\psi$. Classically,

$$\partial_\mu j^{\mu 5} = 0 \tag{19.1}$$

for zero-mass fermions. This equation of motion is true not only in free fermion theory but also, as a classical field equation, in massless QED and QCD.

However, in this chapter, we will see that the true picture is not so simple. We will show that, in gauge theories, the conservation of the axial vector current is actually incompatible with gauge invariance, and that radiative corrections in gauge theories supply a nonzero operator that appears on the right-hand side of Eq. (19.1). This new conservation equation for the axial current has a number of remarkable consequences, which we will discuss in Sections 19.3 and 19.4.

19.1 The Axial Current in Two Dimensions

Eventually, we will want to analyze the current conservation equation for the axial current in massless QCD. However, this discussion will involve some technical complication, so we will first study the physics that violates axial current conservation in a context in which the calculations are relatively simple. A particularly simple model problem is that of two-dimensional massless QED.

The Lagrangian of two-dimensional QED is

$$\mathcal{L} = \bar{\psi}(i\not{D})\psi - \tfrac{1}{4}(F_{\mu\nu})^2, \tag{19.2}$$

with $\mu, \nu = 0, 1$ and $D_\mu = \partial_\mu + ieA_\mu$. The Dirac matrices must be chosen to satisfy the Dirac algebra

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}. \tag{19.3}$$

In two dimensions, this set of relations can be represented by 2×2 matrices; we choose

$$\gamma^0 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \gamma^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \tag{19.4}$$

The Dirac spinors will be two-component fields.

The product of the Dirac matrices, which anticommutes with each of the γ^μ, is

$$\gamma^5 = \gamma^0\gamma^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{19.5}$$

Then, just as in four dimensions, there are two possible currents,

$$j^\mu = \bar{\psi}\gamma^\mu\psi, \qquad j^{\mu 5} = \bar{\psi}\gamma^\mu\gamma^5\psi, \tag{19.6}$$

and both are conserved if there is no mass term in the Lagrangian.

To make the conservation laws quite explicit, we label the components of the fermion field ψ in this spinor basis as

$$\psi = \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix}. \tag{19.7}$$

The subscript indicates the γ^5 eigenvalue. Then, using the explicit representations (19.4) and (19.7), we can rewrite the fermionic part of (19.2) as

$$\mathcal{L} = \psi_+^\dagger i(D_0 + D_1)\psi_+ + \psi_-^\dagger i(D_0 - D_1)\psi_-. \tag{19.8}$$

In the free theory, the field equation of ψ_+ would be

$$i(\partial_0 + \partial_1)\psi_+ = 0; \tag{19.9}$$

the solutions to this equation are waves that move to the right in the one-dimensional space at the speed of light. We will thus refer to the particles associated with ψ_+ as *right-moving* fermions. The quanta associated with ψ_- are, similarly, *left-moving*. This distinction is analogous to the distinction between left- and right-handed particles which gives the physical interpretation of γ^5 in four dimensions. Since the Lagrangian (19.8) contains no terms that mix left- and right-moving fields, it seems obvious that the number currents for these fields are separately conserved. Thus,

$$\partial_\mu\Big(\bar{\psi}\gamma^\mu\Big(\frac{1-\gamma^5}{2}\Big)\psi\Big) = 0, \qquad \partial_\mu\Big(\bar{\psi}\gamma^\mu\Big(\frac{1+\gamma^5}{2}\Big)\psi\Big) = 0. \tag{19.10}$$

It is a curious property of two-dimensional spacetime that the vector and axial vector fermionic currents are not independent of each other. Let $\epsilon^{\mu\nu}$ be

the totally antisymmetric symbol in two dimensions, with $\epsilon^{01} = +1$. Then the two-dimensional Dirac matrices obey the identity

$$\gamma^\mu \gamma^5 = -\epsilon^{\mu\nu}\gamma_\nu. \tag{19.11}$$

The currents $j^{\mu 5}$ and j^μ have the same relation. Thus we can study the properties of the axial vector current by using results that we have already derived for the vector current.

Vacuum Polarization Diagrams

In Section 7.5, we computed the lowest-order vacuum polarization of QED in dimensional regularization. In the limit of zero mass, we found, in Eq. (7.90),

$$i\Pi^{\mu\nu}(q) = -i(q^2 g^{\mu\nu} - q^\mu q^\nu)\frac{2e^2}{(4\pi)^{d/2}}\,\mathrm{tr}[1]\int_0^1 dx\, x(1-x)\frac{\Gamma(2-\frac{d}{2})}{(-x(1-x)q^2)^{2-d/2}}, \tag{19.12}$$

where $\mathrm{tr}[1] = 4$ gives the convention for tracing over Dirac matrices given in Eq. (7.88). If we set $\mathrm{tr}[1] = 2$ to be consistent with (19.4) and then set $d = 2$ in (19.12), we find the finite and well-defined result

$$\begin{aligned} i\Pi^{\mu\nu}(q) &= i(q^2 g^{\mu\nu} - q^\mu q^\nu)\frac{2e^2}{4\pi}\cdot 2\cdot\frac{1}{q^2} \\ &= i\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\Big)\frac{e^2}{\pi}. \end{aligned} \tag{19.13}$$

Notice that this expression has the structure of a photon mass term; the photon receives the mass

$$m_\gamma^2 = \frac{e^2}{\pi}. \tag{19.14}$$

Schwinger showed that this result is exact, and that the photon of two-dimensional QED is a free massive boson.* In the discussion below Eq. (7.72), we pointed out that it is not possible for a vacuum polarization amplitude consistent with the Ward identity to generate a mass for the photon unless it also contains a pole at $q^2 = 0$. In two dimensions, such a pole can arise from the infrared behavior of the fermion-antifermion intermediate state, and we see this behavior explicitly in (19.13).

Once we have an explicit expression for the vacuum polarization, we can find the expectation value of the current induced by a background electromagnetic field. This quantity is generated by the diagram of Fig. 19.1, which gives

$$\int d^2x\, e^{iq\cdot x}\,\langle j^\mu(x)\rangle = \frac{i}{e}\big(i\Pi^{\mu\nu}(q)\big)A_\nu(q) = -\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\Big)\cdot\frac{e}{\pi}A_\nu(q), \tag{19.15}$$

*J. Schwinger, *Phys. Rev.* **128**, 2425 (1962).

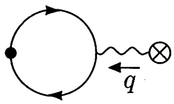


Figure 19.1. Computation of $\langle j^\mu \rangle$ in a background electromagnetic field.

where $A_\nu(q)$ is the Fourier transform of the background field. This quantity manifestly satisfies the current conservation relation $q_\mu \langle j^\mu(q) \rangle = 0$.

The identity (19.11) between the vector and axial vector currents allows us to derive from (19.15) the corresponding expectation value of $j^{\mu 5}$. We find

$$\begin{aligned}\langle j^{\mu 5}(q) \rangle &= -\epsilon^{\mu\nu} \langle j_\nu(q) \rangle \\ &= \epsilon^{\mu\nu} \frac{e}{\pi} \Big(A_\nu(q) - \frac{q_\nu q^\lambda}{q^2} A_\lambda(q) \Big).\end{aligned} \tag{19.16}$$

If the axial vector current were conserved, this object would satisfy the Ward identity. Instead, we find

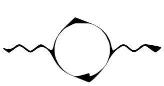
$$q_\mu \langle j^{\mu 5}(q) \rangle = \frac{e}{\pi} \epsilon^{\mu\nu} q_\mu A_\nu(q). \tag{19.17}$$

This is the Fourier transform of the field equation

$$\partial_\mu j^{\mu 5} = \frac{e}{2\pi} \epsilon^{\mu\nu} F_{\mu\nu}. \tag{19.18}$$

Apparently, the axial vector current is not conserved in the presence of electromagnetic fields, as the result of an anomalous behavior of its vacuum polarization diagram.

How could this happen? The Feynman diagrams formally satisfy the Ward identity both for the vector and for the axial vector current. The problem must come in the regularization of the vacuum polarization diagram. By dimensional analysis, we know that this diagram has the form



$$= ie^2 \Big(A g^{\mu\nu} - B \frac{q^\mu q^\nu}{q^2} \Big). \tag{19.19}$$

The coefficient B is a finite integral, and is, in any event, unambiguously determined by the low-energy structure of the theory since it is the residue of the pole in q^2. However, the integral A is logarithmically divergent, so its value depends on the regularization. Dimensional regularization automatically subtracts this integral to set $A = B$; then the vector current Ward identity is satisfied. But then we are led directly to (19.17). We could, alternatively, regularize the integral A so that $A = 0$. Working through the steps of the previous paragraph with this modification, we now find $q_\mu \langle j^{\mu 5}(q) \rangle = 0$, but

$$q_\mu \langle j^\mu(q) \rangle = \frac{e}{\pi} q^\nu A_\nu(q). \tag{19.20}$$

Though the result (19.17) is unpleasant, the result (19.20) would be a complete disaster, since it depends on the unphysical gauge degrees of freedom

of the vector potential. We conclude that it is not possible to regularize two-dimensional QED so that, simultaneously, the theory is gauge invariant and the axial vector current is conserved. The price of requiring gauge invariance is the anomalous nonconservation of the axial current shown in (19.18).

The Axial Vector Current Operator Equation

To understand what happened to the axial current from another viewpoint, we now study the operator equation for the divergence of $j^{\mu 5}$. Varying the Lagrangian (19.2), we find the following equations of motion for the fermion fields:

$$\partial\!\!\!/ \psi = -ie A\!\!\!/ \psi, \qquad \partial_\mu \bar{\psi}\gamma^\mu = ie\bar{\psi}A\!\!\!/. \tag{19.21}$$

By using these equations of motion in the most straightforward way, it is easy to conclude that $\partial_\mu j^{\mu 5} = 0$. However, a closer look at these manipulations reveals some subtleties, which alter the final conclusion.

The axial vector current is a composite operator built out of fermion fields. In the previous chapter we saw that products of local operators are often singular, so we will define the current by placing the two fermion fields at distinct points separated by a distance ϵ and then carefully taking the limit as the two fields approach each other. Explicitly, we define

$$j^{\mu 5} = \underset{\epsilon\to 0}{\text{symm lim}} \left\{ \bar{\psi}(x+\tfrac{\epsilon}{2})\gamma^\mu\gamma^5 \exp\Big[-ie\int_{x-\epsilon/2}^{x+\epsilon/2} dz\cdot A(z)\Big]\psi(x-\tfrac{\epsilon}{2})\right\}. \tag{19.22}$$

Notice that, because we have placed ψ and $\bar{\psi}$ at different points, we must introduce a Wilson line (15.53) in order that the operator be locally gauge invariant. To give $j^{\mu 5}$ the correct transformation properties under Lorentz transformations, the limit $\epsilon \to 0$ should be taken symmetrically,

$$\underset{\epsilon\to 0}{\text{symm lim}} \left\{\frac{\epsilon^\mu}{\epsilon^2}\right\} = 0, \qquad \underset{\epsilon\to 0}{\text{symm lim}} \left\{\frac{\epsilon^\mu \epsilon^\nu}{\epsilon^2}\right\} = \frac{1}{d} g^{\mu\nu}, \tag{19.23}$$

with $d = 2$ in this case.

We now compute the divergence of the axial current defined as in (19.22):

$$\begin{aligned}\partial_\mu j^{\mu 5} = \underset{\epsilon\to 0}{\text{symm lim}} \Bigg\{ & (\partial_\mu\bar{\psi}(x+\tfrac{\epsilon}{2}))\gamma^\mu\gamma^5 \exp\Big[-ie\int_{x-\epsilon/2}^{x+\epsilon/2} dz\cdot A(z)\Big]\psi(x-\tfrac{\epsilon}{2}) \\ & + \bar{\psi}(x+\tfrac{\epsilon}{2})\gamma^\mu\gamma^5 \exp\Big[-ie\int_{x-\epsilon/2}^{x+\epsilon/2} dz\cdot A(z)\Big](\partial_\mu\psi(x-\tfrac{\epsilon}{2})) \\ & + \bar{\psi}(x+\tfrac{\epsilon}{2})\gamma^\mu\gamma^5\big[-ie\epsilon^\nu\partial_\mu A_\nu(x)\big]\psi(x-\tfrac{\epsilon}{2})\Bigg\}. \end{aligned} \tag{19.24}$$

Using the equations of motion (19.21), and keeping terms up to order ϵ, we can reduce this to

$$\begin{aligned}\partial_\mu j^{\mu 5} &= \operatorname*{symm\,lim}_{\epsilon\to 0}\Big\{\bar\psi(x+\tfrac{\epsilon}{2})\big[ie\not{\epsilon}\!\!A(x+\tfrac{\epsilon}{2}) - ie\not{\epsilon}\!\!A(x-\tfrac{\epsilon}{2}) \\ &\qquad - ie\epsilon^\nu\gamma^\mu\partial_\mu A_\nu(x)\big]\gamma^5\psi(x-\tfrac{\epsilon}{2})\Big\} \\ &= \operatorname*{symm\,lim}_{\epsilon\to 0}\Big\{\bar\psi(x+\tfrac{\epsilon}{2})\big[-ie\gamma^\mu\epsilon^\nu(\partial_\mu A_\nu - \partial_\nu A_\mu)\big]\gamma^5\psi(x-\tfrac{\epsilon}{2})\Big\}.\end{aligned} \tag{19.25}$$

Expression (19.25) seems to vanish in the limit $\epsilon \to 0$. However, we must take account of the fact that the product of the fermion operators is singular. In two dimensions, the contraction of fermion fields is

$$\begin{aligned}\psi(y)\bar\psi(z) &= \int \frac{d^2k}{(2\pi)^2} e^{-ik\cdot(y-z)} \frac{i\not{k}}{k^2} \\ &= -\not{\partial}\Big(\frac{i}{4\pi}\log(y-z)^2\Big) \\ &= \frac{-i}{2\pi}\frac{\gamma^\alpha(y-z)_\alpha}{(y-z)^2}.\end{aligned} \tag{19.26}$$

Thus,

$$\bar\psi(x+\tfrac{\epsilon}{2})\Gamma\psi(x-\tfrac{\epsilon}{2}) = \frac{-i}{2\pi}\operatorname{tr}\Big[\frac{\gamma^\alpha\epsilon_\alpha}{\epsilon^2}\Gamma\Big]. \tag{19.27}$$

Notice that the result (19.27) contains an extra minus sign from the interchange of fermion operators.

Because the contraction of fermion fields is singular as $\epsilon \to 0$, the terms of order ϵ in the last line of (19.25) can give a finite contribution. Taking the contraction according to (19.27), we find

$$\partial_\mu j^{\mu 5} = \operatorname*{symm\,lim}_{\epsilon\to 0}\Big\{\frac{-i}{2\pi}\operatorname{tr}\Big[\frac{\gamma^\alpha\epsilon_\alpha}{\epsilon^2}\gamma^\mu\gamma^5\Big]\cdot\big(-ie\epsilon^\nu F_{\mu\nu}\big)\Big\}. \tag{19.28}$$

In two dimensions, $\operatorname{tr}[\gamma^\alpha\gamma^\mu\gamma^5] = 2\epsilon^{\alpha\mu}$. Thus,

$$\partial_\mu j^{\mu 5} = \frac{e}{2\pi}\operatorname*{symm\,lim}_{\epsilon\to 0}\Big\{2\frac{\epsilon_\mu\epsilon^\nu}{\epsilon^2}\Big\}\epsilon^{\mu\alpha}F_{\nu\alpha}. \tag{19.29}$$

Now take the symmetric limit according to the prescription (19.23). We find precisely the anomalous nonconservation equation (19.18). In this derivation, (19.18) appears as an operator relation, rather than in a simple matrix element. Notice that, as in our first derivation of this equation, the assumption of local gauge invariance played a crucial role. If we had defined the axial vector current by reversing the sign of the Wilson line in (19.22), a prescription that would have done violence to local gauge invariance, we would have found the various contributions canceling on the right-hand side of (19.29).

An Example with Fermion Number Nonconservation

To complete our discussion of the two-dimensional axial vector current, we will show that the nonconservation equation (19.18) also has a global aspect. In free fermion theory, the integral of the axial current conservation law gives

$$\int d^2x\, \partial_\mu j^{\mu 5} = N_R - N_L = 0. \tag{19.30}$$

This relation implies that the difference in the number of right-moving and left-moving fermions cannot be changed in any possible process. Combining this with the conservation law for the vector current, we conclude that the number of each type of fermion is separately conserved. From (19.8), we might conclude that these separate conservation laws hold also in two-dimensional QED. However, we have already found that we must be careful in making statements about the axial current.

In two-dimensional QED, the conservation equation for the axial current is replaced by the anomalous nonconservation equation (19.18). If the right-hand side of this equation were the total derivative of a quantity falling off sufficiently rapidly at infinity, its integral would vanish and we would still retain the global conservation law. In fact, $\epsilon^{\mu\nu}F_{\mu\nu}$ is a total derivative:

$$\epsilon^{\mu\nu}F_{\mu\nu} = 2\partial_\mu\big(\epsilon^{\mu\nu}A_\nu\big). \tag{19.31}$$

However, it is easy to imagine examples where the integral of this quantity does not vanish, for example, a world with a constant background electric field. In such a world, the conservation law (19.30) must be violated. But how can this happen?

Let us analyze this problem by thinking about fermions in one space dimension in a background A^1 field that is constant in space and has a very slow time dependence. We will assume that the system has a finite length L, with periodic boundary conditions. Notice that the constant A^1 field cannot be removed by a gauge transformation that satisfies the periodic boundary conditions. One way to see this is to note that the system gives a nonzero value to the Wilson line

$$\exp\Big[-ie\int_0^L dx\, A_1(x)\Big], \tag{19.32}$$

which forms a gauge-invariant closed loop due to the periodic boundary conditions.

Following the derivation of the three-dimensional Hamiltonian, Eq. (3.84), we find that the Hamiltonian of this one-dimensional system is

$$H = \int dx\, \psi^\dagger\big(-i\alpha^1 D_1\big)\psi, \tag{19.33}$$

where $\alpha = \gamma^0\gamma^1 = \gamma^5$. In the components (19.7),

$$H = \int dx \Big\{ -i\psi_+^\dagger(\partial_1 - ieA^1)\psi_+ + i\psi_-^\dagger(\partial_1 - ieA^1)\psi_- \Big\}. \qquad (19.34)$$

For a constant A^1 field, it is easy to diagonalize this Hamiltonian. The eigenstates of the covariant derivatives are wavefunctions

$$e^{ik_n x}, \qquad \text{with} \quad k_n = \frac{2\pi n}{L}, \; n = -\infty, \dots, \infty, \qquad (19.35)$$

to satisfy the periodic boundary conditions. Then the single-particle eigenstates of H have energies

$$\begin{aligned} \psi_+ : \qquad & E_n = +(k_n - eA^1), \\ \psi_- : \qquad & E_n = -(k_n - eA^1). \end{aligned} \qquad (19.36)$$

Each type of fermion has an infinite tower of equally spaced levels. To find the ground state of H, we fill the negative energy levels and interpret holes created among these filled states as antiparticles.

Now adiabiatically change the value of A^1. The fermion energy levels slowly shift in accord with the relations (19.36). If A^1 changes by the finite amount

$$\Delta A^1 = \frac{2\pi}{eL}, \qquad (19.37)$$

which brings the Wilson loop (19.32) back to its original value, the spectrum of H returns to its original form. In this process, each level of ψ_+ moves down to the next position, and each level of ψ_- moves up to the next position, as shown in Fig. 19.2. The occupation numbers of levels should be maintained in this adiabatic process. Thus, remarkably, one right-moving fermion disappears from the vacuum and one extra left-moving fermion appears. At the same time,

$$\begin{aligned} \int d^2x \left(\frac{e}{\pi}\epsilon^{\mu\nu}F_{\mu\nu}\right) &= \int dt\, dx \frac{e}{\pi}\partial_0 A_1 \\ &= \frac{e}{\pi}L(-\Delta A^1) \\ &= -2, \end{aligned} \qquad (19.38)$$

where we have inserted (19.37) in the last line. Thus the integrated form of the anomalous nonconservation equation (19.18) is indeed satisfied:

$$N_R - N_L = \int d^2x \left(\frac{e}{2\pi}\epsilon^{\mu\nu}F_{\mu\nu}\right). \qquad (19.39)$$

Even in this simple example, we see that it is not possible to escape the question of ultraviolet regularization in analyzing the chiral conservation law. Right-moving fermions are lost and left-moving fermions appear from the depths of the fermionic spectrum, $E \to -\infty$. In computing the changes in the separate fermion numbers, we have assumed that the vacuum cannot change the charge it contains at large negative energies. This prescription is gauge invariant, but it leads to the nonconservation of the axial vector current.

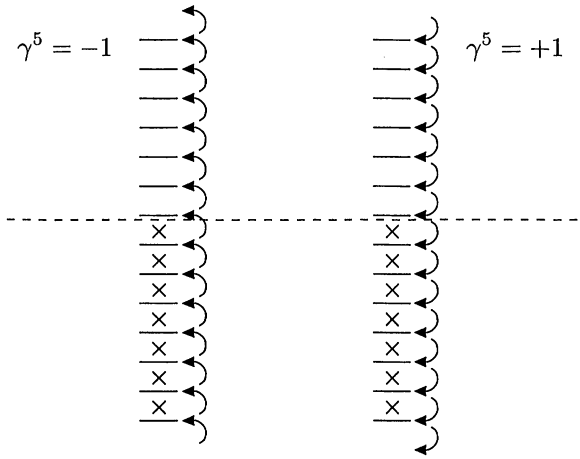


Figure 19.2. Effect on the vacuum state of the Hamiltonian H of one-dimensional QED due to an adiabatic change in the background A^1 field.

19.2 The Axial Current in Four Dimensions

All of the derivations we have just given for the two-dimensional axial current have analogues in four dimensions. In Eq. (3.40), we showed that, in the case of massless fermions, the four-dimensional Dirac equation splits neatly into separate equations for left- and right-handed fermions. If we couple the Dirac equation to a gauge field, we replace derivatives by covariant derivatives. This does not seem to affect the manifest separation between the two helicity components. Thus it seems clear that both the vector and axial vector currents should remain conserved. However, after the analysis we have just completed for the two-dimensional case, we know that we should not take these conservation laws for granted. We will now make a more careful analysis of the axial vector conservation law in four dimensions.

The Axial Vector Current Operator Equation

We begin with the case of massless four-dimensional QED. Of the three arguments that we gave in the previous section for the two-dimensional axial current conservation law, the operator derivation generalizes most easily. The fermion field equations (19.21) are identical in the four-dimensional case. We can again adopt the gauge-invariant definition of the axial vector current (19.22). When we take the divergence of this current, all of the manipulations leading to Eq. (19.25) are still correct.

From this point, we must compute the singular terms in the operator product of the two fermion fields in the limit $\epsilon \to 0$. As in two dimensions, the leading term is given by contracting the two operators using a free-field

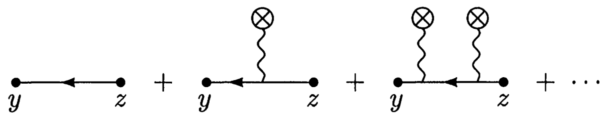


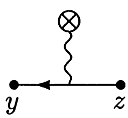
Figure 19.3. Expansion of $\psi(y)\bar{\psi}(z)$ in the presence of a background gauge field.

propagator. This contribution gives

$$\begin{aligned}
\overline{\psi(y)\bar{\psi}}(z) &= \int \frac{d^4k}{(2\pi)^4} e^{-ik\cdot(y-z)} \frac{i\not{k}}{k^2} \\
&= -\not{\partial}\Big(\frac{i}{4\pi^2}\frac{1}{(y-z)^2}\Big) \\
&= \frac{-i}{2\pi^2}\frac{\gamma^\alpha(y-z)_\alpha}{(y-z)^4}.
\end{aligned} \tag{19.40}$$

This is highly singular as $(y-z)\to 0$, but it gives zero when traced with $\gamma^\mu\gamma^5$. To find a nonzero result, we must consider terms of higher order in the expansion of the product of operators.

In a nonzero background gauge field, the contraction of fermion fields is given by the series of diagrams shown in Fig. 19.3. We have computed the leading term in this series in (19.40). The higher terms give less singular terms as $(y-z)\to 0$. The second term in the series is given by



$$= \int \frac{d^4k}{(2\pi)^4}\frac{d^4p}{(2\pi)^4} e^{-i(k+p)\cdot y} e^{ik\cdot z} \frac{i(\not{k}+\not{p})}{(k+p)^2}\big(-ie\not{A}(p)\big)\frac{i\not{k}}{k^2}. \tag{19.41}$$

This contribution leads to

$$\begin{aligned}
&\big\langle \bar{\psi}(x+\tfrac{\epsilon}{2})\gamma^\mu\gamma^5\psi(x-\tfrac{\epsilon}{2})\big\rangle \\
&\quad= \int \frac{d^4k}{(2\pi)^4}\frac{d^4p}{(2\pi)^4} e^{ik\cdot\epsilon} e^{-ip\cdot x}\, \mathrm{tr}\Big[\big(-\gamma^\mu\gamma^5\big)\frac{i(\not{k}+\not{p})}{(k+p)^2}\big(-ie\not{A}(p)\big)\frac{i\not{k}}{k^2}\Big] \\
&\quad= \int \frac{d^4k}{(2\pi)^4}\frac{d^4p}{(2\pi)^4} e^{ik\cdot\epsilon} e^{-ip\cdot x}\frac{4e\epsilon^{\mu\alpha\beta\gamma}(k+p)_\alpha A_\beta(p)k_\gamma}{k^2(k+p)^2}.
\end{aligned} \tag{19.42}$$

To evaluate the limit $\epsilon\to 0$, we can expand the integrand for large k. Then

$$\begin{aligned}
\big\langle \bar{\psi}(x+\tfrac{\epsilon}{2})\gamma^\mu\gamma^5\psi(x-\tfrac{\epsilon}{2})\big\rangle &\sim 4e\epsilon^{\mu\alpha\beta\gamma}\int \frac{d^4p}{(2\pi)^4} e^{-ip\cdot x} p_\alpha A_\beta(p) \int \frac{d^4k}{(2\pi)^4} e^{ik\cdot\epsilon}\frac{k_\gamma}{k^4} \\
&= 4e\epsilon^{\alpha\beta\mu\gamma}\big(\partial_\alpha A_\beta(x)\big)\frac{\partial}{\partial\epsilon^\gamma}\Big(\frac{i}{16\pi^2}\log\frac{1}{\epsilon^2}\Big) \\
&= 2e\epsilon^{\alpha\beta\mu\gamma}F_{\alpha\beta}(x)\Big(\frac{-i}{8\pi^2}\frac{\epsilon_\gamma}{\epsilon^2}\Big).
\end{aligned} \tag{19.43}$$

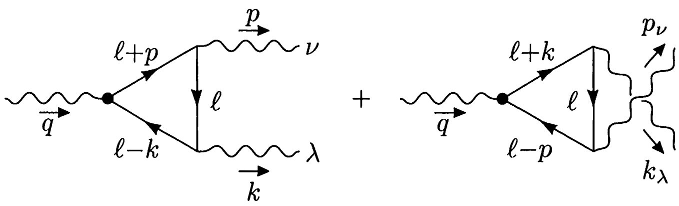


Figure 19.4. Diagrams contributing to the two-photon matrix element of the divergence of the axial vector current.

Substituting this expression into (19.25), we find

$$\partial_\mu j^{\mu 5} = \operatorname{symm}\lim_{\epsilon\to 0}\left\{\frac{e}{4\pi^2}\epsilon^{\alpha\beta\mu\gamma}F_{\alpha\beta}\Big(\frac{-i\epsilon_\gamma}{\epsilon^2}\Big)\big(-ie\epsilon^\nu F_{\mu\nu}\big)\right\}. \tag{19.44}$$

Now take the symmetric limit $\epsilon \to 0$ in four dimensions. We find

$$\partial_\mu j^{\mu 5} = -\frac{e^2}{16\pi^2}\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu}. \tag{19.45}$$

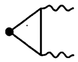
This equation, which expresses the anomalous nonconservation of the four-dimensional axial current, is known as the Adler-Bell-Jackiw anomaly. Adler and Bardeen proved that this operator relation is actually correct to all orders in QED perturbation theory and receives no further radiative corrections.†

Triangle Diagrams

We can confirm the Adler-Bell-Jackiw relation by checking, in standard perturbation theory, that the divergence of the axial vector current has a nonzero matrix element to create two photons. To do this, we must analyze the matrix element

$$\int d^4x e^{-iq\cdot x}\,\langle p,k|\, j^{\mu 5}(x)\,|0\rangle = (2\pi)^4\delta^{(4)}(p+k-q)\epsilon^*_\nu(p)\epsilon^*_\lambda(k)\mathcal{M}^{\mu\nu\lambda}(p,k). \tag{19.46}$$

The leading-order diagrams contributing to $\mathcal{M}^{\mu\nu\lambda}$ are shown in Fig. 19.4. The first diagram gives the contribution



$$= (-1)(-ie)^2\int\frac{d^4\ell}{(2\pi)^4}\operatorname{tr}\left[\gamma^\mu\gamma^5\frac{i(\not\ell-\not k)}{(\ell-k)^2}\gamma^\lambda\frac{i\not\ell}{\ell^2}\gamma^\nu\frac{i(\not\ell+\not p)}{(\ell+p)^2}\right], \tag{19.47}$$

and the second diagram gives an identical contribution with (p,ν) and (k,λ) interchanged.

It is easy to give a formal argument that the matrix element of the divergence of the axial current vanishes at this order. Taking the divergence of the axial current in (19.46) is equivalent to dotting this quantity with iq_μ.

†S. Adler and W. A. Bardeen, *Phys. Rev.* **182**, 1517 (1969); S. Adler, in Deser, et. al. (1970).

Now we operate on the right-hand side of (19.47) as we do to prove a Ward identity. Replace

$$q_\mu \gamma^\mu \gamma^5 = (\not{\ell} + \not{p} - \not{\ell} + \not{k})\gamma^5 = (\not{\ell} + \not{p})\gamma^5 + \gamma^5(\not{\ell} - \not{k}). \tag{19.48}$$

Each momentum factor combines with the numerator adjacent to it to cancel the corresponding denominator. This brings (19.47) into the form

$iq_\mu \cdot$ 

$$= e^2 \int \frac{d^4\ell}{(2\pi)^4} \operatorname{tr}\left[\gamma^5 \frac{(\not{\ell} - \not{k})}{(\ell - k)^2}\gamma^\lambda \frac{\not{\ell}}{\ell^2}\gamma^\nu + \gamma^5 \gamma^\lambda \frac{\not{\ell}}{\ell^2}\gamma^\nu \frac{(\not{\ell} + \not{p})}{(\ell + p)^2}\right]. \tag{19.49}$$

Now pass γ^ν through γ^5 in the second term and shift the integral over the first term according to $\ell \to (\ell + k)$:

$iq_\mu \cdot$ 

$$= e^2 \int \frac{d^4\ell}{(2\pi)^4} \operatorname{tr}\left[\gamma^5 \frac{\not{\ell}}{\ell^2}\gamma^\lambda \frac{(\not{\ell} + \not{k})}{(\ell + k)^2}\gamma^\nu - \gamma^5 \frac{\not{\ell}}{\ell^2}\gamma^\nu \frac{(\not{\ell} + \not{p})}{(\ell + p)^2}\gamma^\lambda\right]. \tag{19.50}$$

This expression is manifestly antisymmetric under the interchange of (p, ν) and (k, λ), so the contribution of the second diagram in Fig. 19.4 precisely cancels (19.47).

However, because this derivation involves a shift of the integration variable, we should look closely at whether this shift is allowed by the regularization. From (19.47), we see that the integral that must be shifted is divergent. If the diagram is regulated with a simple momentum cutoff, or even with Pauli-Villars regularization, it turns out that the shift leaves over a finite, nonzero term. In Chapter 7, we encountered a similar problem in our discussion of the QED vacuum polarization diagram. We evaded the problem there by using dimensional regularization. Dimensional regularization of the diagrams of Fig. 19.4 will automatically insure the validity of the QED Ward identities for the photon emission vertices,

$$p_\nu \mathcal{M}^{\mu\nu\lambda} = k_\lambda \mathcal{M}^{\mu\nu\lambda} = 0. \tag{19.51}$$

But in the analysis of the axial vector current, even dimensional regularization has an extra subtlety, because γ^5 is an intrinsically four-dimensional object. In their original paper on dimensional regularization,‡ 't Hooft and Veltman suggested using the definition

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 \tag{19.52}$$

in d dimensions. This definition has the consequence that γ^5 anticommutes with γ^μ for $\mu = 0, 1, 2, 3$ but commutes with γ^μ for other values of μ.

In the evaluation of (19.47), the external indices and the momenta p, k, q all live in the physical four dimensions, but the loop momentum ℓ has components in all dimensions. Write

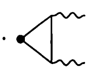
$$\ell = \ell_\parallel + \ell_\perp, \tag{19.53}$$

‡G. 't Hooft and M. J. G. Veltman, *Nucl. Phys.* **B44**, 189 (1972).

where the first term has nonzero components in dimensions $0, 1, 2, 3$ and the second term has nonzero components in the other $d-4$ dimensions. Because γ^5 commutes with the γ^μ in these extra dimensions, identity (19.48) is modified to

$$q_\mu \gamma^\mu \gamma^5 = (\not{\ell} + \not{k})\gamma^5 + \gamma^5(\not{\ell} - \not{p}) - 2\gamma^5 \not{\ell}_\perp . \tag{19.54}$$

The first two terms cancel according to the argument given above; the shift in (19.50) is justified by the dimensional regularization. However, the third term of (19.54) gives an additional contribution:

$iq_\mu \cdot$ 

$$= e^2 \int \frac{d^4\ell}{(2\pi)^4} \operatorname{tr}\left[-2\gamma^5 \not{\ell}_\perp \frac{(\not{\ell} - \not{k})}{(\ell - k)^2} \gamma^\lambda \frac{\not{\ell}}{\ell^2} \gamma^\nu \frac{(\not{\ell} + \not{p})}{(\ell + p)^2} \right]. \tag{19.55}$$

To evaluate this contribution, combine denominators in the standard way, and shift the integration variable $\ell \to \ell + P$, where $P = xk - yp$. In expanding the numerator, we must retain one factor each of γ^ν, γ^λ, $\not{p}$, and $\not{k}$ to give a nonzero trace with γ^5. This leaves over one factor of $\not{\ell}_\perp$ and one factor of $\not{\ell}$ which must also be evaluated with components in extra dimensions in order to give a nonzero integral. The factors $\not{\ell}_\perp$ anticommute with the other Dirac matrices in the problem and thus can be moved to adjacent positions. Then we must evaluate the integral

$$\int \frac{d^4\ell}{(2\pi)^4} \frac{\not{\ell}_\perp \not{\ell}_\perp}{(\ell^2 - \Delta)^3}, \tag{19.56}$$

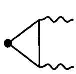
where Δ is a function of k, p, and the Feynman parameters. Using

$$(\not{\ell}_\perp)^2 = \ell_\perp^2 \to \frac{(d-4)}{d} \ell^2 \tag{19.57}$$

under the symmetrical integration, we can evaluate (19.56) as

$$\frac{i}{(4\pi)^{d/2}} \frac{(d-4)}{2} \frac{\Gamma(2-\frac{d}{2})}{\Gamma(3)\Delta^{2-d/2}} \underset{d\to 4}{\longrightarrow} \frac{-i}{2(4\pi)^2}. \tag{19.58}$$

Notice the behavior in which a logarithmically divergent integral contributes a factor $(d-4)$ in the denominator and allows an anomalous term, formally proportional to $(d-4)$, to give a finite contribution. The remainder of the algebra in the evaluation of (19.55) is straightforward. The terms involving the momentum shift P cancel, and we find

$iq_\mu \cdot$ 

$$\begin{aligned} &= e^2 \left(\frac{-i}{2(4\pi)^2} \right) \operatorname{tr}\left[2\gamma^5 (-\not{k}) \gamma^\lambda \not{p} \gamma^\nu \right] \\ &= \frac{e^2}{4\pi^2} \epsilon^{\alpha\lambda\beta\nu} k_\alpha p_\beta . \end{aligned} \tag{19.59}$$

This term is symmetric under the interchange of (p,ν) with (k,λ), so the second diagram of Fig. 19.4 gives an equal contribution. Thus,

$$\begin{aligned}\langle p,k|\,\partial_\mu j^{\mu 5}(0)\,|0\rangle &= -\frac{e^2}{2\pi^2}\epsilon^{\alpha\nu\beta\lambda}(-ip_\alpha)\epsilon^*_\nu(p)(-ik_\beta)\epsilon^*_\lambda(k)\\ &= -\frac{e^2}{16\pi^2}\,\langle p,k|\,\epsilon^{\alpha\nu\beta\lambda}F_{\alpha\nu}F_{\beta\lambda}(0)\,|0\rangle\,,\end{aligned} \tag{19.60}$$

as we would expect from the Adler-Bell-Jackiw anomaly equation.

Chiral Transformation of the Functional Integral

A third way of understanding the Adler-Bell-Jackiw anomaly comes from analyzing the conservation law for the axial vector current from the functional integral for the fermion field. In Section 9.6, we used the functional integral to derive the current conservation equations and the Ward identities associated with any symmetry of the Lagrangian. It is instructive to see how this argument breaks down when we apply it to the chiral symmetry of massless fermions.

We first review the standard derivation of the axial vector Ward identities following the method of Section 9.6. Starting from the fermionic functional integral

$$Z = \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\exp\Big[i\int d^4x\,\bar{\psi}(i\not{D})\psi\Big], \tag{19.61}$$

make the change of variables

$$\begin{aligned}\psi(x) &\to \psi'(x) = (1+i\alpha(x)\gamma^5)\psi(x),\\ \bar{\psi}(x) &\to \bar{\psi}'(x) = \bar{\psi}(1+i\alpha(x)\gamma^5).\end{aligned} \tag{19.62}$$

Since the global chiral rotation, with constant α, is a symmetry of the Lagrangian, the only new terms in the Lagrangian that result from (19.62) contain derivatives of α. Thus,

$$\begin{aligned}\int d^4x\,\bar{\psi}'(i\not{D})\psi' &= \int d^4x\big[\bar{\psi}(i\not{D})\psi - \partial_\mu\alpha(x)\bar{\psi}\gamma^\mu\gamma^5\psi\big]\\ &= \int d^4x\big[\bar{\psi}(i\not{D})\psi + \alpha(x)\partial_\mu(\bar{\psi}\gamma^\mu\gamma^5\psi)\big].\end{aligned} \tag{19.63}$$

Then, by varying the Lagrangian with respect to $\alpha(x)$, we derive the classical conservation equation for the axial current. By carrying out a similar manipulation on the functional expression for a correlation function, as in Eq. (9.102), we would derive the associated Ward identities.

In the argument just given, we assumed that the functional measure does not change when we change variables from $\psi'(x)$ to ψ. This seems reasonable, because the relation of ψ' and ψ in (19.62) looks like a unitary transformation. However, we should examine this point more closely.* First, we must carefully

*K. Fujikawa, *Phys. Rev. Lett.* **42**, 1195 (1979); *Phys. Rev.* **D21**, 2848 (1980).

define the functional measure. To do this, expand the fermion field in a basis of eigenstates of $\not{D}$. Define right and left eigenvectors of $\not{D}$ by

$$(i\not{D})\phi_m = \lambda_m \phi_m, \qquad \hat{\phi}_m(i\not{D}) = -iD_\mu \hat{\phi}_m \gamma^\mu = \lambda_m \hat{\phi}_m. \tag{19.64}$$

For zero background A_μ field, these eigenstates are Dirac wavefunctions of definite momentum; the eigenvalues satisfy

$$\lambda_m^2 = k^2 = (k^0)^2 - (\mathbf{k})^2. \tag{19.65}$$

For a fixed A_μ field, this is also the asymptotic form of the eigenvalues for large k. These eigenfunctions give us a basis that we can use to expand ψ and $\bar{\psi}$:

$$\psi(x) = \sum_m a_m \phi_m(x), \qquad \bar{\psi}(x) = \sum_m \hat{a}_m \hat{\phi}_m(x), \tag{19.66}$$

where a_m, $\hat{a}_m$ are anticommuting coefficients multiplying the c-number eigenfunctions (19.64). The functional measure over ψ, $\bar{\psi}$ can then be defined as

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} = \prod_m da_m d\hat{a}_m, \tag{19.67}$$

and the functional measure over ψ', $\bar{\psi}'$ can be defined in the same way.

If $\psi'(x) = (1 + i\alpha(x)\gamma^5)\psi(x)$, the expansion coefficients of ψ and ψ' are related by a infinitesimal linear transformation $(1 + C)$, computed as follows:

$$a'_m = \sum_n \int d^4x \phi_m^\dagger(x)(1 + i\alpha(x)\gamma^5)\phi_n(x) a_n = \sum_n (\delta_{mn} + C_{mn}) a_n. \tag{19.68}$$

Then

$$\mathcal{D}\psi'\mathcal{D}\bar{\psi}' = \mathcal{J}^{-2} \cdot \mathcal{D}\psi\mathcal{D}\bar{\psi}, \tag{19.69}$$

where $\mathcal{J}$ is the Jacobian determinant of the transformation $(1 + C)$. The inverse of $\mathcal{J}$ appears in (19.69) as a result of the rule (9.63) or (9.69) for fermionic integration. To evaluate $\mathcal{J}$, we write

$$\mathcal{J} = \det(1 + C) = \exp\big[\operatorname{tr}\log(1 + C)\big] = \exp\Big[\sum_n C_{nn} + \cdots\Big], \tag{19.70}$$

and we can ignore higher order terms in the last line because C is infinitesimal. Thus,

$$\log \mathcal{J} = i \int d^4x\, \alpha(x) \sum_n \phi_n^\dagger(x)\gamma^5\phi_n(x). \tag{19.71}$$

The coefficient of $\alpha(x)$ looks like $\operatorname{tr}[\gamma^5] = 0$. However, we must regularize the sum over eigenstates n in a gauge-invariant way. The natural choice is

$$\sum_n \phi_n^\dagger(x)\gamma^5\phi_n(x) = \lim_{M\to\infty} \sum_n \phi_n^\dagger(x)\gamma^5\phi_n(x) e^{\lambda_n^2/M^2}. \tag{19.72}$$

As (19.65) indicates, the sign of λ_n^2 will be negative at large momentum after a Wick rotation; thus, the sign in the exponent of the convergence factor is given correctly. We can write (19.72) in an operator form

$$\begin{aligned}\sum_n \phi_n^\dagger(x)\gamma^5\phi_n(x) &= \lim_{M\to\infty}\sum_n \phi_n^\dagger(x)\gamma^5 e^{(i\not{D})^2/M^2}\phi_n(x) \\ &= \lim_{M\to\infty}\langle x|\,\mathrm{tr}\left[\gamma^5 e^{(i\not{D})^2/M^2}\right]|x\rangle\,,\end{aligned} \tag{19.73}$$

where, in the second line, we trace over Dirac indices.

To evaluate (19.73), we rewrite $(i\not{D})^2$ according to (16.107). In our present conventions, this equation reads

$$(i\not{D})^2 = -D^2 + \frac{e}{2}\sigma^{\mu\nu}F_{\mu\nu}, \tag{19.74}$$

with $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu,\gamma^\nu]$. Since we are taking the limit $M \to \infty$, we can concentrate our attention on the asymptotic part of the spectrum, where the momentum k is large and we can expand in powers of the gauge field. To obtain a nonzero trace with γ^5, we must bring down four Dirac matrices from the exponent. The leading term is given by expanding the exponent to order $(\sigma\cdot F)^2$, and then ignoring the background A_μ field in all other terms. This gives

$$\begin{aligned}&\lim_{M\to\infty}\langle x|\,\mathrm{tr}\left[\gamma^5 e^{(-D^2+(e/2)\sigma\cdot F)/M^2}\right]|x\rangle \\ &= \lim_{M\to\infty}\mathrm{tr}\left[\gamma^5\frac{1}{2!}\left(\frac{e}{2M^2}\sigma^{\mu\nu}F_{\mu\nu}(x)\right)^2\right]\langle x|\,e^{-\partial^2/M^2}|x\rangle\,.\end{aligned} \tag{19.75}$$

The matrix element in (19.75) can be evaluated by a Wick rotation:

$$\begin{aligned}\langle x|\,e^{-\partial^2/M^2}|x\rangle &= \lim_{x\to y}\int\frac{d^4k}{(2\pi)^4}e^{-ik\cdot(x-y)}e^{k^2/M^2} \\ &= i\int\frac{d^4k_E}{(2\pi)^4}e^{-k_E^2/M^2} \\ &= i\frac{M^4}{16\pi^2}.\end{aligned} \tag{19.76}$$

Then (19.75) reduces to

$$\begin{aligned}&\lim_{M\to\infty}\frac{-ie^2}{8\cdot 16\pi^2}M^4\,\mathrm{tr}\left[\gamma^5\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma\frac{1}{(M^2)^2}F_{\mu\nu}F_{\lambda\sigma}(x)\right] \\ &= -\frac{e^2}{32\pi^2}\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu}(x).\end{aligned} \tag{19.77}$$

Thus,

$$\mathcal{J} = \exp\left[-i\int d^4x\,\alpha(x)\left(\frac{e^2}{32\pi^2}\epsilon^{\mu\nu\lambda\sigma}F_{\mu\nu}F_{\lambda\sigma}(x)\right)\right]. \tag{19.78}$$

In all, we find that, after the change of variables (19.62), the functional integral (19.61) takes the form

$$Z = \int \mathcal{D}\psi\mathcal{D}\bar{\psi} \exp\left[i \int d^4x \left(\bar{\psi}(i\not{D})\psi + \alpha(x)\left\{ \partial_\mu j^{\mu 5} + \frac{e^2}{16\pi^2}\epsilon^{\mu\nu\lambda\sigma} F_{\mu\nu}F_{\lambda\sigma} \right\} \right) \right]. \tag{19.79}$$

Varying the exponent with respect to $\alpha(x)$, we find precisely the Adler-Bell-Jackiw anomaly equation.

This derivation of the axial vector anomaly is especially interesting because it generalizes readily to any even dimensionality. The functional derivation always picks out for the right-hand side of the anomaly equation the pseudoscalar operator built from gauge fields that has the same dimension, d, as the divergence of the current. In two dimensions, this derivation leads immediately to (19.18). As long as d is even, we can always construct a matrix γ^5 that anticommutes with all of the Dirac matrices by taking their product. Then, the functional derivation leads straightforwardly to the result

$$\partial_\mu j^{\mu 5} = (-1)^{n+1} \frac{2e^n}{n!(4\pi)^n} \epsilon^{\mu_1\mu_2\cdots\mu_{2n}} F_{\mu_1\mu_2} \cdots F_{\mu_{2n-1}\mu_{2n}}, \tag{19.80}$$

where $n = d/2$.

At the end of the previous section, we argued that the axial vector anomaly leads to global nonconservation of fermionic charges in a two-dimensional system with a macroscopic electric field. In the same way, the four-dimensional anomaly equation leads to global nonconservation of the number of left- and right-handed fermions in background fields in which the right-hand side of (19.45) is nonzero. These are field configurations with parallel electric and magnetic fields. In Problem 19.1, we work out an example of four-dimensional massless fermions in a simple situation of this type and show that the fermion numbers are indeed violated, in a manner similar to what we saw at the end of Section 19.1, in accord with the Adler-Bell-Jackiw anomaly.

19.3 Goldstone Bosons and Chiral Symmetries in QCD

The Adler-Bell-Jackiw anomaly has a number of important implications for QCD. To describe these, we must first discuss the chiral symmetries of QCD systematically. In this discussion, we will ignore all but the lightest quarks u and d. In many analyses of the low-energy structure of the strong interactions, one also treats the s quark as light; this gives results that naturally generalize the ones we will find below.

The fermionic part of the QCD Lagrangian is

$$\mathcal{L} = \bar{u}i\not{D}u + \bar{d}i\not{D}d - m_u\bar{u}u - m_d\bar{d}d. \tag{19.81}$$

If the u and d quarks are very light, the last two terms are small and can be neglected. Let us study the implications of making this approximation. If

we ignore the u and d masses, the Lagrangian (19.81) of course has isospin symmetry, the symmetry of an $SU(2)$ unitary transformation mixing the u and d fields. However, because the classical Lagrangian for massless fermions contains no coupling between left- and right-handed quarks, this Lagrangian actually is symmetric under the separate unitary transformations

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \to U_L \begin{pmatrix} u \\ d \end{pmatrix}_L, \qquad \begin{pmatrix} u \\ d \end{pmatrix}_R \to U_R \begin{pmatrix} u \\ d \end{pmatrix}_R. \tag{19.82}$$

It is useful to separate the $U(1)$ and $SU(2)$ parts of these transformations; then the symmetry group of the classical, massless QCD Lagrangian is $SU(2) \times SU(2) \times U(1) \times U(1)$. Let Q denote the quark doublet, with chiral components

$$Q_L = \left(\frac{1-\gamma^5}{2}\right)\begin{pmatrix} u \\ d \end{pmatrix}, \qquad Q_R = \left(\frac{1+\gamma^5}{2}\right)\begin{pmatrix} u \\ d \end{pmatrix}. \tag{19.83}$$

Then we can write the currents associated with these symmetries as

$$\begin{aligned} j_L^\mu &= \bar{Q}_L \gamma^\mu Q_L, \qquad & j_R^\mu &= \bar{Q}_R \gamma^\mu Q_R, \\ j_L^{\mu a} &= \bar{Q}_L \gamma^\mu \tau^a Q_L, \qquad & j_R^{\mu a} &= \bar{Q}_R \gamma^\mu \tau^a Q_R, \end{aligned} \tag{19.84}$$

where $\tau^a = \sigma^a/2$ represent the generators of $SU(2)$. The sums of left- and right-handed currents give the baryon number and isospin currents

$$j^\mu = \bar{Q}\gamma^\mu Q, \qquad j^{\mu a} = \bar{Q}\gamma^\mu \tau^a Q. \tag{19.85}$$

The corresponding symmetries are the transformations (19.82) with $U_L = U_R$. The differences of the currents (19.84) give the corresponding axial vector currents $j^{\mu 5}$, $j^{\mu 5a}$:

$$j^{\mu 5} = \bar{Q}\gamma^\mu \gamma^5 Q, \qquad j^{\mu 5a} = \bar{Q}\gamma^\mu \gamma^5 \tau^a Q. \tag{19.86}$$

In the discussion to follow, we will derive conclusions about the strong interactions by assuming that the classical conservation laws for these currents are not spoiled by anomalies. We will show below that this assumption is correct for the isotriplet currents $j^{\mu 5a}$ but not for $j^{\mu 5}$.

The vector $SU(2) \times U(1)$ transformations are manifest symmetries of the strong interactions, and the associated currents lead to familiar conservation laws. What about the orthogonal, axial vector, transformations? These do not correspond to any obvious symmetry of the strong interactions. In 1960, Nambu and Jona-Lasinio hypothesized that these are accurate symmetries of the strong interactions that are spontaneously broken.[†] This idea has led to a correct and surprisingly detailed description of the properties of the strong interactions at low energy.

[†] Y. Nambu and G. Jona-Lasinio, *Phys. Rev.* **122**, 345 (1961).

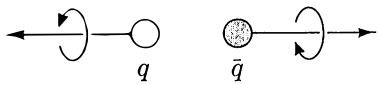


Figure 19.5. A quark-antiquark pair with zero total momentum and angular momentum.

Spontaneous Breaking of Chiral Symmetry

Before we describe the consequences of spontaneously broken chiral symmetry, let us ask why we might expect the chiral symmetries to be spontaneously broken in the first place. In the theory of superconductivity, a small electron-electron attraction leads to the appearance of a condensate of electron pairs in the ground state of a metal. In QCD, quarks and antiquarks have strong attractive interactions, and, if these quarks are massless, the energy cost of creating an extra quark-antiquark pair is small. Thus we expect that the vacuum of QCD will contain a condensate of quark-antiquark pairs. These fermion pairs must have zero total momentum and angular momentum. Thus, as Fig. 19.5 shows, they must contain net chiral charge, pairing left-handed quarks with the antiparticles of right-handed quarks. The vacuum state with a quark pair condensate is characterized by a nonzero vacuum expectation value for the scalar operator

$$\langle 0|\, \overline{Q}Q\, |0\rangle = \langle 0|\, \overline{Q}_L Q_R + \overline{Q}_R Q_L\, |0\rangle \neq 0, \tag{19.87}$$

which transforms under (19.82) with $U_L \neq U_R$. The expectation value signals that the vacuum mixes the two quark helicities. This allows the u and d quarks to acquire effective masses as they move through the vacuum. Inside quark-antiquark bound states, the u and d quarks would appear to move as if they had a sizable effective mass, even if they had zero mass in the original QCD Lagrangian.

The vacuum expectation value (19.87) signals the spontaneous breaking of the full symmetry group (19.82) down to the subgroup of vector symmetries with $U_L = U_R$. Thus there are four spontaneously broken continuous symmetries, associated with the four axial vector currents. At the end of Section 11.1, we proved Goldstone's theorem, which states that every spontaneously broken continuous symmetry of a quantum field theory leads to a massless particle with the quantum numbers of a local symmetry rotation. This means that, in QCD with massless u and d quarks, we should find four spin-zero particles with the correct quantum numbers to be created by the four axial vector currents.

The real strong interactions do not contain any massless particles, but they do contain an isospin triplet of relatively light mesons, the pions. These particles are known to have odd parity (as we expect if they are quark-antiquark bound states). Thus, they can be created by the axial isospin currents. We can parametrize the matrix element of $j^{\mu 5a}$ between the vacuum

and an on-shell pion by writing

$$\langle 0| \, j^{\mu 5a}(x) \, |\pi^b(p)\rangle = -ip^\mu f_\pi \delta^{ab} e^{-ip\cdot x}, \tag{19.88}$$

where a, b are isospin indices and f_π is a constant with the dimensions of $(\text{mass})^1$. We show in Problem 19.2 that the value of f_π can be determined from the rate of π^+ decay through the weak interaction; one finds $f_\pi = 93$ MeV. For this reason, f_π is often called the *pion decay constant.* If we contract (19.88) with p_μ and use the conservation of the axial currents, we find that an on-shell pion must satisfy $p^2 = 0$, that is, it must be massless, as required by Goldstone's theorem.

If we now restore the quark mass terms in (19.81), the axial currents are no longer exactly conserved. The equation of motion of the quark field is now

$$i\not{D}Q = \mathbf{m}Q, \qquad -iD_\mu \bar{Q}\gamma^\mu = \bar{Q}\mathbf{m}, \tag{19.89}$$

where

$$\mathbf{m} = \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix} \tag{19.90}$$

is the quark mass matrix. Then one can readily compute

$$\partial_\mu j^{\mu 5a} = i\bar{Q}\{\mathbf{m}, \tau^a\}\gamma^5 Q. \tag{19.91}$$

Using this equation together with (19.88), we find

$$\langle 0| \, \partial_\mu j^{\mu 5a}(0) \, |\pi^b(p)\rangle = -p^2 f_\pi \delta^{ab} = \langle 0| \, i\bar{Q}\{\mathbf{m}, \tau^a\}\gamma^5 Q \, |\pi^b(p)\rangle . \tag{19.92}$$

The last expression is an invariant quantity times

$$\operatorname{tr}\big[\{\mathbf{m}, \tau^a\}\tau^b\big] = \tfrac{1}{2}\delta^{ab}(m_u + m_d). \tag{19.93}$$

Thus, the quark mass terms give the pions masses of the form

$$m_\pi^2 = (m_u + m_d)\frac{M^2}{f_\pi}. \tag{19.94}$$

The mass parameter M has been estimated to be of order 400 MeV. Thus, to give the observed pion mass of 140 MeV, one needs only $(m_u + m_d) \sim 10$ MeV. This is a small perturbation on the strong interactions.

This argument has an interesting implication for the nature of the isospin symmetry of the strong interactions. In the limit in which the u and d quarks have zero mass in the Lagrangian, these quarks acquire large, equal effective masses from the vacuum with spontaneously broken chiral symmetry. As long as the masses m_u and m_d in the Lagrangian are small compared to the effective mass, the u and d quarks will behave inside hadrons as though they are approximately degenerate. Thus the isospin symmetry of the strong interactions need have nothing to do with a fundamental symmetry linking u and d; it follows for any arbitrary relation between m_u and m_d, provided that both of these parameters are much less than 300 MeV. Similarly, the approximate $SU(3)$ symmetry of the strong interactions follows if the fundamental mass of the s quark is also small compared to the strong interaction scale. The best

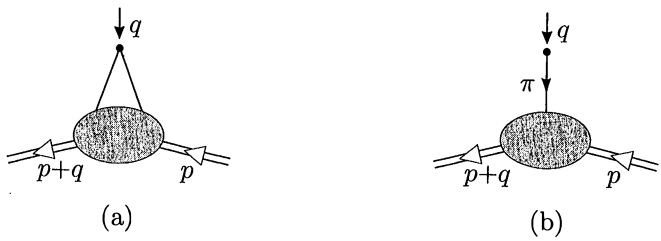


Figure 19.6. Matrix element of the axial isospin current in the nucleon: (a) kinematics of the amplitude; (b) contribution that leads to a pole in q^2.

current estimates of the mass ratios $m_u : m_d : m_s$ are in fact $1 : 2 : 40$, so that the fundamental Lagrangian of the strong interactions shows no sign of flavor symmetry among the quark masses.‡

The identification of the pions as Goldstone bosons of broken chiral symmetry has a number of implications for hadronic matrix elements. Here we will give only one example. In the following argument, we will work in the limit of exact chiral symmetry, ignoring the small corrections from the u and d masses.

The matrix element of the axial isospin current in the nucleon, a quantity that enters the theory of neutron and nuclear β decay, can be written in terms of form factors as follows:

$$\langle N|\, j^{\mu 5a}(q)\,|N\rangle = \bar{u}\left[\gamma^\mu\gamma^5 F_1^5(q^2) + \frac{i\sigma^{\mu\nu}q_\nu}{2m}\gamma^5 F_2^5(q^2) + q^\mu\gamma^5 F_3^5(q^2)\right]\tau^a u. \tag{19.95}$$

The kinematics of the vertex is shown in Fig. 19.6. Notice that there is one more possible form factor than in the vector case, Eq. (6.33). The value of F_1^5 at $q^2 = 0$ is not restricted by the value of any manifestly conserved charge. Conventionally, one writes simply

$$F_1^5(0) = g_A. \tag{19.96}$$

However, we will now show that the value of this quantity can be computed.

If we ignore quark masses, the axial vector current in (19.95) is conserved, so the form factors satisfy

$$\begin{aligned} 0 &= \bar{u}(p')\left[\not{q}\gamma^5 F_1^5(q^2) + q^2\gamma^5 F_3^5(q^2)\right]u(p) \\ &= \bar{u}(p')\left[(\not{p}' - \not{p})\gamma^5 F_1^5(q^2) + q^2\gamma^5 F_3^5(q^2)\right]u(p) \\ &= \bar{u}(p')\left[2m_N\gamma^5 F_1^5(q^2) + q^2\gamma^5 F_3^5(q^2)\right]u(p). \end{aligned} \tag{19.97}$$

‡The determination of the fundamental quark masses is reviewed by J. Gasser and H. Leutwyler, *Phys. Repts.* **87**, 77 (1982).

Thus, we find

$$g_A = \lim_{q^2\to 0} \frac{q^2}{2m_N} F_3^5(q^2). \tag{19.98}$$

This equation implies that $g_A = 0$ unless F_3^5 contains a pole in q^2. Such a pole would imply the presence of a physical massless particle, but fortunately, there is one available—the massless pion. The process in which the current creates a pion that is then absorbed by the nucleon indeed leads to a pole in $F_3^5(q^2)$, as shown in Fig. 19.6(b).

Let us now compute this pole term and use it to determine g_A. The low-energy pion-nucleon interaction is conventionally parametrized by the Lagrangian

$$\Delta\mathcal{L} = ig_{\pi NN}\pi^a \overline{N}\gamma^5\sigma^a N. \tag{19.99}$$

The amplitude for the current $j^{\mu 5a}$ to create the pion is given by (19.88). Then the contribution of Fig. 19.6(b) to the current vertex is

$$-g_{\pi NN}\bar{u}(2\tau^a\gamma^5)u \cdot \frac{i}{q^2} \cdot (iq^\mu f_\pi). \tag{19.100}$$

Thus,

$$F_3^5(q^2) = \frac{1}{q^2} \cdot 2f_\pi g_{\pi NN}. \tag{19.101}$$

We find that g_A is given by a combination of f_π, the nucleon mass, and the pion-nucleon coupling constant:

$$g_A = \frac{f_\pi}{m_N} g_{\pi NN}. \tag{19.102}$$

This strange identity, called the *Goldberger-Treiman relation*, is satisfied experimentally to 5% accuracy.

The identification of the pion as the Goldstone boson of spontaneously broken chiral symmetry leads to numerous other predictions for current matrix elements and pion scattering amplitudes. In particular, the leading terms of the pion-pion and pion-nucleon scattering amplitudes at low energy can be computed directly in terms of f_π by arguments similar to one just given.*

Anomalies of Chiral Currents

Up to this point, we have discussed the chiral symmetries of QCD according to the classical current conservation equations. We must now ask whether these equations are affected by the Adler-Bell-Jackiw anomaly, and what the consequences of that modification are.

To begin, we study the modification of the chiral conservation laws due to the coupling of the quark currents to the gluon fields of QCD. The arguments given in the previous section go through equally well in the case of

*The detailed consequences of spontaneously broken chiral symmetry are worked out in a very clear manner in Georgi (1984).

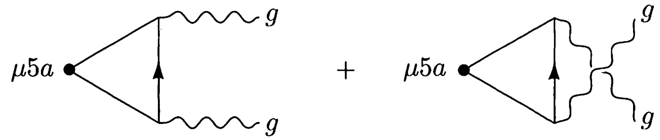


Figure 19.7. Diagrams that lead to an axial vector anomaly for a chiral current in QCD.

massless fermions coupling to a non-Abelian gauge field, so we expect that an axial vector current will receive an anomalous contribution from the diagrams shown in Fig. 19.7. The anomaly equation should be the Abelian result, supplemented by an appropriate group theory factor. In addition, since the axial current is gauge invariant, the anomaly must also be gauge invariant. That is, it must contain the full non-Abelian field strength, including its nonlinear terms. These terms are actually included in the functional derivation of the anomaly given at the end of Section 19.2.

For the axial currents of QCD, written in (19.86), we can read the group theory factors for the Adler-Bell-Jackiw anomaly from the diagrams of Fig. 19.7. For the axial isospin currents,

$$\partial_\mu j^{\mu 5a} = -\frac{g^2}{16\pi^2}\epsilon^{\alpha\beta\mu\nu}F^c_{\alpha\beta}F^d_{\mu\nu}\cdot \operatorname{tr}\left[\tau^a t^c t^d\right], \tag{19.103}$$

where $F^c_{\mu\nu}$ is a gluon field strength, τ^a is an isospin matrix, t^c is a color matrix, and the trace is taken over colors and flavors. In this case, we find

$$\operatorname{tr}\left[\tau^a t^c t^d\right] = \operatorname{tr}[\tau^a]\operatorname{tr}[t^c t^d] = 0, \tag{19.104}$$

since the trace of a single τ^a vanishes. Thus the conservation of the axial isospin currents is unaffected by the Adler-Bell-Jackiw anomaly of QCD. However, in the case of the isospin singlet axial current, the matrix τ^a is replaced by the matrix 1 on flavors, and we find

$$\partial_\mu j^{\mu 5} = -\frac{g^2 n_f}{32\pi^2}\epsilon^{\alpha\beta\mu\nu}F^c_{\alpha\beta}F^c_{\mu\nu}, \tag{19.105}$$

where n_f is the number of flavors; $n_f = 2$ in our current model.

Thus, the isospin singlet axial current is not in fact conserved in QCD. The divergence of this current is equal to a gluon operator with nontrivial matrix elements between hadron states. Some subtle questions remain concerning the effects of this operator. In particular, it can be shown, as we saw for the two-dimensional axial anomaly in Eq. (19.31), that the right-hand side of (19.105) is a total divergence. Nevertheless, again in accord with our experience in two dimensions, there are physically reasonable field configurations in which the four-dimensional integral of this term takes a nonzero value. This topic is discussed further at the end of Section 22.3. In any event, Eq. (19.105) indeed implies that QCD has no isosinglet axial symmetry and no associated Goldstone boson. This equation explains why the strong interactions contain

no light isosinglet pseudoscalar meson with mass comparable to that of the pions.

Though the axial isospin currents have no axial anomaly from QCD interactions, they do have an anomaly associated with the coupling of quarks to electromagnetism. Again referring to the diagrams of Fig. 19.7, we see that the electromagnetic anomaly of the axial isospin currents is given by

$$\partial_\mu j^{\mu 5a} = -\frac{e^2}{16\pi^2}\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu}\cdot \mathrm{tr}\big[\tau^a Q^2\big], \tag{19.106}$$

where $F_{\mu\nu}$ is the electromagnetic field strength, Q is the matrix of quark electric charges,

$$Q = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & -\frac{1}{3} \end{pmatrix}, \tag{19.107}$$

and the trace again runs over flavors and colors. Since the matrices in the trace do not depend on color, the color sum simply gives a factor of 3. The flavor trace is nonzero only for $a = 3$; in that case, the electromagnetic anomaly is

$$\partial_\mu j^{\mu 53} = -\frac{e^2}{32\pi^2}\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu}. \tag{19.108}$$

Because the current $j^{\mu 53}$ annihilates a π^0 meson, Eq. (19.108) indicates that the axial vector anomaly contributes to the matrix element for the decay $\pi^0 \to 2\gamma$. We will now show that, in fact, it gives the leading contribution to this amplitude. Again, we work in the limit of massless u and d quarks, so that the chiral symmetries are exact up to the effects of the anomaly.

Consider the matrix element of the axial current between the vacuum and a two-photon state:

$$\langle p,k|\, j^{\mu 53}(q)\,|0\rangle = \epsilon^*_\nu\epsilon^*_\lambda \mathcal{M}^{\mu\nu\lambda}(p,k). \tag{19.109}$$

This is the same matrix element (19.46) that we studied in QED perturbation theory in Section 19.2. Now, however, we will study the general properties of this matrix element by expanding it in form factors. In general, the amplitude can be decomposed by writing all possible tensor structures and applying the restrictions that follow from symmetry under the interchange of (p,ν) and (k,λ) and the QED Ward identities (19.51). This leaves three possible structures:

$$\begin{aligned}\mathcal{M}^{\mu\nu\lambda} &= q^\mu\epsilon^{\nu\lambda\alpha\beta}p_\alpha k_\beta\mathcal{M}_1 + \big(\epsilon^{\mu\nu\alpha\beta}k^\lambda - \epsilon^{\mu\lambda\alpha\beta}p^\nu\big)k_\alpha p_\beta\mathcal{M}_2 \\ &\quad + \Big[\big(\epsilon^{\mu\nu\alpha\beta}p^\lambda - \epsilon^{\mu\lambda\alpha\beta}k^\nu\big)k_\alpha p_\beta - \epsilon^{\mu\nu\lambda\sigma}(p-k)_\sigma p\cdot k\Big]\mathcal{M}_3.\end{aligned} \tag{19.110}$$

The second term satisfies (19.51) by virtue of the on-shell conditions $p^2 = k^2 = 0$.

Now contract (19.110) with (iq_μ) to take the divergence of the axial vector current. We find

$$iq_\mu\mathcal{M}^{\mu\nu\lambda} = iq^2\epsilon^{\nu\lambda\alpha\beta}p_\alpha k_\beta\mathcal{M}_1 - i\epsilon^{\mu\nu\lambda\sigma}q_\mu(p-k)_\sigma p\cdot k\mathcal{M}_3; \tag{19.111}$$

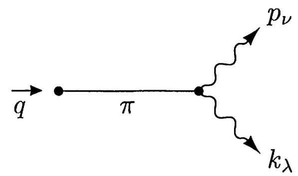


Figure 19.8. Contribution that leads to a pole in the axial vector current form factor $\mathcal{M}_1$.

the other terms automatically give zero. Using $q = p + k$, $q^2 = 2p \cdot k$, we can simplify this to

$$iq_\mu \mathcal{M}^{\mu\nu\lambda} = iq^2 \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta (\mathcal{M}_1 + \mathcal{M}_3). \tag{19.112}$$

The whole quantity is proportional to q^2 and apparently vanishes in the limit $q^2 \to 0$. This contrasts with the prediction of the axial vector anomaly. Taking the matrix element of the right-hand side of (19.108), we find

$$iq_\mu \mathcal{M}^{\mu\nu\lambda} = -\frac{e^2}{4\pi^2} \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta. \tag{19.113}$$

The conflict can be resolved if one of the form factors appearing in (19.112) contains a pole in q^2. Such a pole can arise through the process shown in Fig. 19.8, in which the current creates a π^0 meson which subsequently decays to two photons. The amplitude for the current to create the meson is given by (19.88). Let us parametrize the pion decay amplitude as

$$i\mathcal{M}(\pi^0 \to 2\gamma) = iA\epsilon^*_\nu \epsilon^*_\lambda \epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta, \tag{19.114}$$

where A is a constant to be determined. Then the contribution of the process of Fig. 19.8 to the amplitude $\mathcal{M}^{\mu\nu\lambda}$ defined in (19.109) is

$$\big(iq^\mu f_\pi\big) \frac{i}{q^2} \big(iA\epsilon^{\nu\lambda\alpha\beta} p_\alpha k_\beta\big). \tag{19.115}$$

This is a contribution to the form factor $\mathcal{M}_1$,

$$\mathcal{M}_1 = \frac{-i}{q^2} f_\pi \cdot A, \tag{19.116}$$

plus terms regular at $q^2 = 0$. Now, by equating (19.112) to (19.113), we determine A in terms of the coefficient of the anomaly:

$$A = \frac{e^2}{4\pi^2} \frac{1}{f_\pi}. \tag{19.117}$$

From the decay matrix element (19.114), it is straightforward to work out the decay rate of the π^0. Note that, though we have worked out the decay matrix element in the limit of a massless π^0, we must supply the physically

correct kinematics which depends on the π^0 mass. Including a factor 1/2 for the phase space of identical particles, we find

$$\begin{aligned}\Gamma(\pi^0 \to 2\gamma) &= \frac{1}{2m_\pi}\frac{1}{8\pi}\frac{1}{2}\sum_{\text{pols.}}\left|\mathcal{M}(\pi^0 \to 2\gamma)\right|^2 \\ &= \frac{1}{32\pi m_\pi}\cdot A^2\cdot 2(p\cdot k)^2 \\ &= A^2\cdot\frac{m_\pi^3}{64\pi}.\end{aligned} \tag{19.118}$$

Thus, finally,

$$\Gamma(\pi^0 \to 2\gamma) = \frac{\alpha^2}{64\pi^3}\frac{m_\pi^3}{f_\pi^2}. \tag{19.119}$$

This relation, which provides a direct measure of the coefficient of the Adler-Bell-Jackiw anomaly, is satisfied experimentally to an accuracy of a few percent.

19.4 Chiral Anomalies and Chiral Gauge Theories

Up to this point, we have coupled gauge fields to fermions in a parity-symmetric manner, replacing the derivative in the Dirac equation by a covariant derivative. This procedure couples the gauge field to the vector current of fermions. However, this procedure gives only a subset of the possible couplings of fermions to gauge bosons. In this section we will construct more general, parity-asymmetric, couplings and discuss their interplay with the axial vector anomaly.

We will focus primarily on theories of massless fermions. If the Lagrangian contains no fermion mass terms, it has no terms that mix the two helicity states of a Dirac fermion. Thus, in a theory that contains massless Dirac fermions ψ_i, we can write the kinetic energy term in the helicity basis (3.36) as

$$\mathcal{L} = \psi_{Li}^\dagger i\bar{\sigma}\cdot\partial\psi_{Li} + \psi_{Ri}^\dagger i\sigma\cdot\partial\psi_{Ri}. \tag{19.120}$$

There is no difficulty in coupling this system to a gauge field by assigning the left-handed fields ψ_{Li} to one representation of the gauge group G and assigning the right-handed fields to a different representation. For example, we might assign the left-handed fields to a representation r of G and take the right-handed fields to be invariant under G. This gives

$$\mathcal{L} = \psi_{Li}^\dagger i\bar{\sigma}\cdot D\psi_{Li} + \psi_{Ri}^\dagger i\sigma\cdot\partial\psi_{Ri}, \tag{19.121}$$

with $D_\mu = \partial_\mu - igA_\mu^a t_r^a$. In more conventional notation, (19.121) becomes

$$\mathcal{L} = \bar{\psi}i\gamma^\mu\Big(\partial_\mu - igA_\mu^a t_r^a\Big(\frac{1-\gamma^5}{2}\Big)\Big)\psi. \tag{19.122}$$

It is straightforward to verify that the classical Lagrangian (19.122) is invariant to the local gauge transformation

$$\begin{aligned} \psi &\to \left(1 + i\alpha^a t^a \left(\frac{1-\gamma^5}{2}\right)\right)\psi, \\ A^a_\mu &\to A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c, \end{aligned} \tag{19.123}$$

which generalizes (15.46). Since the right-handed fields are free fields, we can even eliminate these fields and write a gauge-invariant Lagrangian for purely left-handed fermions.

The idea of gauge fields that couple only to left-handed fermions plays a central role in the construction of a theory of weak interactions. The coupling of the W boson to quarks and leptons described in (17.31) can be derived by assigning the left-handed components of quarks and leptons to doublets of an $SU(2)$ gauge symmetry

$$Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \qquad L_L = \begin{pmatrix} \nu \\ \ell \end{pmatrix}_L, \tag{19.124}$$

and then identifying the W bosons as gauge fields that couple to this $SU(2)$ group. In this picture, it is the restriction of the symmetry to left-handed fields that leads to the helicity structure of the weak interaction effective Lagrangian. We will discuss a complete, explicit model of weak interactions, incorporating this idea, in the next chapter.

To work out the general properties of chirally coupled fermions, it is useful to rewrite their Lagrangian with one further transformation. Below Eq. (3.38), we noted that the quantity $\sigma^2\psi_R^*$ transforms under Lorentz transformations as a left-handed field. Thus it is useful to rewrite the right-handed components in (19.120) as new left-handed fermions, by defining

$$\psi'_{Li} = \sigma^2 \psi^*_{Ri}, \qquad \psi'^\dagger_{Li} = \psi^T_{Ri}\sigma^2. \tag{19.125}$$

This transformation relabels the right-handed fermions as antifermions and calls their left-handed antiparticles a new species of left-handed fermions. By using (3.38), we can rewrite the Lagrangian for the right-handed fermions as

$$\int d^4x\, \psi^\dagger_{Ri} i\sigma \cdot \partial \psi_{Ri} = \int d^4x\, \psi'^\dagger_{Li} i\bar{\sigma} \cdot \partial \psi'_{Li}. \tag{19.126}$$

The minus sign from fermion interchange cancels the minus sign from integration by parts. Notice that, if the fermions are coupled to gauge fields in the representation r, this manipulation changes the covariant derivative as follows:

$$\begin{aligned} \psi^\dagger_R i\sigma \cdot (\partial - igA^a t^a_r)\psi_R &= \psi'^\dagger_L i\bar{\sigma} \cdot (\partial + igA^a (t^a_r)^T)\psi'_L \\ &= \psi'^\dagger_L i\bar{\sigma} \cdot (\partial - igA^a t^a_{\bar{r}})\psi'_L. \end{aligned} \tag{19.127}$$

Thus the new fields ψ'_L belong to the conjugate representation to r, for which the representation matrices are given by (15.82). In this notation, QCD with

n_f flavors of massless fermions is rewritten as an $SU(3)$ gauge theory coupled to n_f massless fermions in the 3 and n_f massless fermions in the $\bar{3}$ representation of $SU(3)$. The most general gauge theory of massless fermions would simply assign left-handed fermions to an arbitrary, reducible representation R of the gauge group G. We have just seen that rewriting a system of Dirac fermions leads to $R = r \oplus \bar{r}$, a *real* representation in the sense described below (15.82). Conversely, if R is not a real representation, then the theory cannot be rewritten in terms of Dirac fermions and is intrinsically chiral.

The rewriting (19.125) transforms the mass term of the QCD Lagrangian as follows:

$$m\bar{\psi}_i\psi_i = m\big(\psi_R^\dagger\psi_L + \text{h.c.}\big) = -m\big(\psi_{Li}'^T\sigma^2\psi_{Li} + \text{h.c.}\big). \tag{19.128}$$

This has the form of the *Majorana* mass term that we encountered in Problem 3.4. The most general mass term that can be built purely from left-handed fermion fields is

$$\Delta\mathcal{L}_M = M_{ij}\psi_{Li}^T\sigma^2\psi_{Lj} + \text{h.c.} \tag{19.129}$$

The matrix M_{ij} is symmetric under $i \leftrightarrow j$, since the minus sign from the antisymmetry of σ^2 is compensated by a minus sign from fermion interchange. This mass term is gauge invariant if M_{ij} is invariant under G. For example, the mass term in (19.128) couples 3 and $\bar{3}$ indices together in an $SU(3)$ singlet combination. In general, a gauge-invariant mass term exists if the representation containing the fermions is *strictly real*, in the sense described below (15.82). In an intrinsically chiral theory, there is no possible gauge-invariant mass term. We will see in the next chapter that, in the gauge theory of the weak interactions, mass terms for the quarks and leptons are forbidden by gauge invariance. We will present a solution to this problem in Section 20.2.

At the classical level, there is no restriction on the representation R of the left-handed fermions. However, at the level of one-loop corrections, many possible choices become inconsistent due to the axial vector anomaly. In a gauge theory of left-handed massless fermions, consider computing the diagrams of Fig. 19.9, in which the external fields are non-Abelian gauge bosons and the marked vertex represents the gauge symmetry current

$$j^{\mu a} = \bar{\psi}\gamma^\mu\Big(\frac{1-\gamma^5}{2}\Big)t^a\psi. \tag{19.130}$$

The gauge boson vertices also contain factors of $(1-\gamma^5)/2$. The three projectors can be moved together into a single factor. Then, if we regularize this diagram as in Section 19.2, the term containing a γ^5 has an axial vector anomaly that leads to the relation

$$\langle p,\nu,b;k,\lambda,c|\,\partial_\mu j^{\mu a}\,|0\rangle = \frac{g^2}{8\pi^2}\epsilon^{\alpha\nu\beta\lambda}p_\alpha k_\beta\cdot\mathcal{A}^{abc}, \tag{19.131}$$

where $\mathcal{A}^{abc}$ is a trace over group matrices in the representation R:

$$\mathcal{A}^{abc} = \text{tr}\big[t^a\{t^b,t^c\}\big]. \tag{19.132}$$

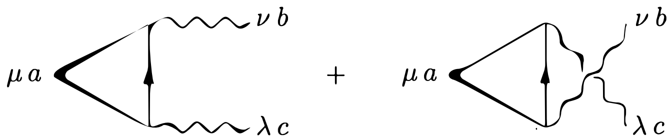


Figure 19.9. Diagrams contributing to the anomaly of a gauge symmetry current in a chiral gauge theory.

This equation implies that, unless $\mathcal{A}^{abc}$ vanishes, the current $j^{\mu a}$ is not conserved. The factor (19.132) is totally symmetric in (a, b, c), so this condition is independent of which current is treated as an external operator. As we described in Sections 19.1 and 19.2, we can change the regularization of the diagram so that the external current is conserved, but only at the price of violating the conservation of one of the other two currents in the diagram.

Since the whole construction of a theory with local gauge invariance is based on the existence of an exact global symmetry, the violation of the conservation of $j^{\mu a}$ does violence to the structure of the theory. For example, triangle diagrams of the form of Fig. 19.9 will now generate divergent gauge boson mass terms and will upset the delicate relations between three- and four-point vertices discussed in Chapter 16. These relations, following from the Ward identity, were necessary to insure the cancellation of unphysical states and the unitarity of the S-matrix. The only way to avoid this problem is to insist that $\mathcal{A}^{abc} = 0$ as a fundamental consistency condition for chiral gauge theories.† Gauge theories satisfying this condition are said to be *anomaly free*.

As an example of the application of this condition, consider the prototype weak interaction gauge theory that we presented in (19.124). If the two gauge bosons in Fig. 19.9 are $SU(2)$ gauge bosons and the current $j^{\mu a}$ is an $SU(2)$ gauge current, we would evaluate (19.132) by substituting $t^a = \tau^a = \sigma^a/2$ and using the relation $\{\sigma^b, \sigma^c\} = 2\delta^{bc}$. This gives

$$\mathcal{A}^{abc} = \frac{1}{8}\,\mathrm{tr}\big[\sigma^a \cdot 2\delta^{bc}\big] = 0, \tag{19.133}$$

so the consistency condition is satisfied. If the fermions in (19.124) also couple to electromagnetism, there is an additional consistency condition that we would find by taking the current in Fig. 19.9 to be the electromagnetic current. The factor $\mathcal{A}^{abc}$ for this case is

$$\mathrm{tr}\big[Q\{\tau^b, \tau^c\}\big], \tag{19.134}$$

where Q is the matrix of electric charges. If we simplify as in (19.133), the trace (19.134) becomes

$$\tfrac{1}{2}\,\mathrm{tr}\big[Q\big]\,\delta^{bc}. \tag{19.135}$$

†D. J. Gross and R. Jackiw, *Phys. Rev.* **D6**, 477 (1972).

This factor is proportional to the sum of the fermion electric charges, which does not vanish either for quarks or for leptons. However, if we sum over one quark doublet and one lepton doublet, with a factor 3 for colors, we find

$$\mathrm{tr}\big[Q\big] = 3 \cdot (\tfrac{2}{3} - \tfrac{1}{3}) + (0 - 1) = 0. \tag{19.136}$$

Remarkably, the weak interaction gauge theory described by (19.124) can be consistently combined with QED only if the theory contains equal numbers of quark and lepton doublets.

We complete this section by working out more generally the condition that a chiral gauge theory be anomaly free. We will first derive some basic properties of the anomaly factor $\mathcal{A}^{abc}$ and then apply these to chiral gauge theories with simple gauge groups.

If the fermion representation R is real, R is equivalent to its conjugate reprsentation $\overline{R}$. Thus, as we described below (15.82), t^a_R is related by a unitary transformation to $t^a_{\overline{R}} = -(t^a_R)^T$. Since (19.132) is invariant to unitary transformations of the t^a, we can replace t^a_R by $t^a_{\overline{R}}$. Then

$$\begin{aligned}\mathcal{A}^{abc} &= \mathrm{tr}\big[(-t^a)^T\big\{(-t^b)^T, (-t^c)^T\big\}\big] \\ &= -\,\mathrm{tr}\big[\big\{t^c, t^b\big\}t^a\big] \\ &= -\mathcal{A}^{abc}.\end{aligned} \tag{19.137}$$

Thus, if R is real, the gauge theory is automatically anomaly free. As a special case, any gauge theory of Dirac fermions is anomaly free.

In more general circumstances, we can simplify the calculation of $\mathcal{A}^{abc}$ by noting that it is an invariant of the gauge group G that is totally symmetric with three indices in the adjoint representation. For some possible groups, a suitable invariant may not exist, and in those cases $\mathcal{A}^{abc}$ must vanish. For example, in $SU(2)$ the adjoint representation has spin 1. The symmetric product of two spin-1 multiplets gives spin 0 plus spin 2, with no spin-1 component. Thus, there is no symmetric tensor coupling two spin-1 indices to give a spin 1. The factor (19.132) must then vanish in any $SU(2)$ gauge theory. We saw this happen in an explicit example in Eq. (19.133).

In $SU(n)$ groups, $n \geq 3$, there is a unique symmetric invariant d^{abc} of the required type. It appears in the anticommutator of representation matrices of the fundamental representation:

$$\big\{t^a_n, t^b_n\big\} = \tfrac{1}{n}\delta^{ab} + d^{abc}t^c_n. \tag{19.138}$$

The uniqueness of this invariant implies that, in an $SU(n)$ gauge theory, any trace of the form of (19.132) is proportional to d^{abc}. For each representation r, we can define an *anomaly coefficient* $A(r)$ by

$$\mathrm{tr}\big[t^a_r\big\{t^b_r, t^c_r\big\}\big] = \tfrac{1}{2}A(r)d^{abc}. \tag{19.139}$$

For the fundamental representation, we can see from (19.138) that $A(n) = 1$. It follows from the argument of (19.137) that

$$A(\bar{r}) = -A(r). \tag{19.140}$$

For higher representations, the anomaly coefficients can be worked out using methods similar to those we used in Section 15.4 to compute $C_2(r)$. For example, we show in Problem 19.3 that, if a and s are the $SU(n)$ representations corresponding to antisymmetric and symmetric two-index tensors, then

$$A(a) = n - 4, \qquad A(s) = n + 4. \tag{19.141}$$

An $SU(n)$ gauge theory is anomaly free if the anomaly coefficients of the various irreducible components of the fermion multiplet R sum to zero. For example, the $SU(n)$ gauge theory of left-handed fermions with representation content

$$R = a + (n-4)\bar{n} \tag{19.142}$$

is anomaly free.

Of the various simple Lie groups listed below (15.72), only $SU(n)$, $SO(4n+2)$, and E_6 have complex representations. Of these, only $SU(n)$ and $SO(6)$, which has the same Lie algebra as $SU(4)$, have a symmetric invariant of the type required to build the anomaly. Gauge theories based on $SO(4n+2)$, $n \geq 2$, and on E_6 are automatically anomaly free. The groups $SO(10)$ and E_6 have been suggested as candidates for the *grand unified* gauge symmetry of particle physics, which we will discuss in Section 22.2.

There is one further constraint on the representation content of a chiral gauge theory, which comes from considering its coupling to gravity. It is possible to show that the diagrams of Fig. 19.9 give an anomaly contribution when computed with a gauge current $j^{\mu a}$ and external gravitational fields. The group-theory factor that multiplies this diagram is

$$\mathrm{tr}\left[t^a_R\right]. \tag{19.143}$$

This factor automatically vanishes if the gauge group is non-Abelian. However, if the gauge group of the theory contains $U(1)$ factors, the theory cannot be consistently coupled to gravity unless each of the $U(1)$ generators is traceless.‡

Once we have constructed a consistent chiral gauge theory, we have an additional problem of finding a prescription for calculating in this theory consistently. In a vector-like gauge theory, we can define ultraviolet-divergent diagrams with dimensional regularization. This guarantees that the divergent diagrams will be regulated in a way that respects the Ward identities of local gauge invariance. To generalize dimensional regularization to chiral gauge theories, we need to introduce a dimensional continuation of γ^5. The 't Hooft-Veltman definition used to define the chiral current in Section 19.2 is not satisfactory, because this definition does not manifestly respect the conservation of the gauge currents. A useful alternative procedure is to define γ^5 formally as an object that anticommutes with all of the γ^μ. This prescription gives unambiguous, gauge-invariant results for amplitudes that are not proportional to $\epsilon^{\mu\nu\lambda\sigma}$, at least through two-loop order. In Section 21.3, we will

‡L. Alvarez-Gaumé and E. Witten, *Nucl. Phys.* **B234**, 269 (1984).

use this prescription to compute loop diagrams in weak interaction theory. As a last resort, one can always compute with a non-gauge-invariant regulator and add non-gauge-invariant counterterms to the theory so that the gauge theory Ward identities remain valid.

19.5 Anomalous Breaking of Scale Invariance

There is one more important example of a symmetry that is an invariance at the classical level and is broken by quantum corrections. This is the classical scale invariance of a massless field theory with a dimensionless coupling constant. In Chapter 12, we saw that a quantum field theory with no classical dimensionful parameters still depends on a mass scale through the regularization of ultraviolet divergences, or, equivalently, through the running of coupling constants. We have already seen how to analyze this induced dependence on the renormalization scale using the Callan-Symanzik equation. In this section, we will show how the violation of classical scale invariance by quantum corrections can be described as a current conservation anomaly.

In this book we have avoided giving a careful treatment of the energy-momentum tensor of a quantum field theory. In Section 2.2, we used Noether's theorem to demonstrate that the invariance of a quantum field theory under spacetime translations implies the presence of a conserved tensor $T^{\mu\nu}$. In Section 9.6, we gave an alternative derivation of this result using the functional integral formalism. However, to discuss the theory of scale invariance, we will need some more detailed properties of the energy-momentum tensor. We will now simply state these properties and refer elsewhere for their derivations.*

The tensor $T^{\mu\nu}$ defined by expressions (2.17) and (9.99) is called the *canonical energy-momentum tensor.* The expressions that defined this tensor do not imply that $T^{\mu\nu}$ is symmetric. In fact, this tensor need not be symmetric, and, in a gauge theory, it need not be gauge-invariant. However, it is always possible to convert $T^{\mu\nu}$ into a symmetric and gauge-invariant tensor $\Theta^{\mu\nu}$ by the addition

$$\Theta^{\mu\nu} = T^{\mu\nu} + \partial_\lambda \Sigma^{\mu\nu\lambda}, \tag{19.144}$$

where $\Sigma^{\mu\nu\lambda}$ is antisymmetric under interchange of μ and λ. The form of the added term implies that $\Theta^{\mu\nu}$ is conserved if $T^{\mu\nu}$ is, and that the global energy-momentum four-vector is unchanged,

$$P^\nu = \int d^3x\, T^{0\nu} = \int d^3x\, \Theta^{0\nu}. \tag{19.145}$$

A scale transformation of a scalar field theory can be defined as a transformation of variables

$$\phi(x) \to e^{-D\sigma}\phi(xe^{-\sigma}), \tag{19.146}$$

*The conclusions presented in the next three paragraphs are derived with care in C. G. Callan, S. Coleman, and R. Jackiw, *Ann. Phys.* **59**, 42 (1970).

with $D = 1$, the canonical mass dimension of the field. The scale transformation is defined similarly in theories of fermion and gauge fields. If this transformation is an invariance of the classical Lagrangian, as it will be if there are no dimensionful couplings, this theory will possess a conserved current D^μ, called the *dilatation current.* The dilatation current has a simple relation to the symmetric energy-momentum tensor $\Theta^{\mu\nu}$: $D^\mu = \Theta^{\mu\nu}x_\nu$, so that

$$\partial_\mu D^\mu = \Theta^\mu{}_\mu. \tag{19.147}$$

The derivation of these results from Noether's theorem is not straightforward. There is a simpler derivation, which, however, uses formalism beyond the scope of this book. If the quantum field theory under consideration is coupled to gravity, then the energy-momentum tensor can be identified as the source of the gravitational field. This energy-momentum tensor can be found by varying the Lagrangian $\mathcal{L}_m$ of matter fields with respect to the spacetime metric $g_{\mu\nu}(x)$. This construction gives a manifestly symmetric and gauge-invariant tensor, which turns out to be $\Theta^{\mu\nu}$:

$$\Theta^{\mu\nu} = 2\frac{\delta}{\delta g_{\mu\nu}(x)}\int d^4x\,\mathcal{L}_m. \tag{19.148}$$

A scale transformation can be represented as a change in the spacetime metric

$$g_{\mu\nu}(x) \to e^{2\sigma} g_{\mu\nu}(x). \tag{19.149}$$

Combining (19.148) and (19.149), we see that the change in the Lagrangian induced by this transformation is just the trace of $\Theta^{\mu\nu}$. This will be equal by Noether's theorem to the divergence of the corresponding current, giving us back Eq. (19.147).

In QED, it is not hard to guess the form of the symmetric energy-momentum tensor:

$$\Theta^{\mu\nu} = -F^{\mu\lambda}F^\nu{}_\lambda + \tfrac{1}{4}g^{\mu\nu}(F_{\lambda\sigma})^2 + \tfrac{1}{2}\bar\psi i(\gamma^\mu D^\nu + \gamma^\nu D^\mu)\psi - g^{\mu\nu}\bar\psi(i\not{D} - m)\psi. \tag{19.150}$$

This is a gauge-invariant symmetric tensor that leads to the familiar expression for the total energy,

$$H = \int d^3x\Big\{\tfrac{1}{2}(E^2 + B^2) + \psi^\dagger(-i\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\nabla} + m)\psi\Big\}. \tag{19.151}$$

For future reference, we note that these results are true at the classical level in any spacetime dimension d. In four dimensions, the trace of the gauge field terms cancels automatically. After using the Dirac equation, which is valid as an operator equation of motion, we find that the trace of $\Theta^{\mu\nu}$ is given by

$$\Theta^\mu{}_\mu = m\bar\psi\psi \tag{19.152}$$

and indeed vanishes, classically, if $m = 0$.

When quantum corrections are included, we know that a scale transformation is not a symmetry of the theory, since the same theory referred to a

larger scale contains a different value of the renormalized coupling constant. The shift of the renormalized coupling is

$$g \to g + \sigma\beta(g), \tag{19.153}$$

and the corresponding change in the Lagrangian is

$$\sigma\beta(g)\frac{\partial}{\partial g}\mathcal{L}. \tag{19.154}$$

Thus, when quantum corrections are included, the equation for the dilatation current in a classically scale-invariant theory should read

$$\partial_\mu D^\mu = \Theta^\mu{}_\mu = \beta(g)\frac{\partial}{\partial g}\mathcal{L}. \tag{19.155}$$

In massless QED, we can write this formula most usefully by rescaling the gauge fields so that the coupling constant e is removed from the covariant derivative: $eA^\mu \to A^\mu$. Then e appears only in the term

$$\mathcal{L} = -\frac{1}{4e^2}(F_{\lambda\sigma})^2 + \cdots, \tag{19.156}$$

and Eq. (19.155) reads

$$\Theta^\mu{}_\mu = \frac{\beta(e)}{2e^3}(F_{\lambda\sigma})^2. \tag{19.157}$$

This relation, which says that the trace of the symmetric energy-momentum tensor takes a nonzero value as a result of quantum corrections, is known as the *trace anomaly.*

We should be able to check the trace anomaly equation (19.157) directly in perturbation theory. We now evaluate the trace of $\Theta^{\mu\nu}$ explicitly to one-loop order. The formalism we have set up is very similar to that of the background field calculation of the β function done in Section 16.6. As in that section, we will integrate over the fluctuating parts of quantum fields in the presence of a background field with a nonzero $F_{\mu\nu}$. Equation (19.157) predicts that this integration will lead to the expression

$$\langle\Theta^\mu{}_\mu\rangle = C\int\frac{d^4k}{(2\pi)^4}A_\mu(-k)(k^2g^{\mu\nu} - k^\mu k^\nu)A_\nu(k), \tag{19.158}$$

where A_μ is the background field and the constant C is equal to $\beta(e)/e^3$.

Since we will be using dimensional regularization, we should begin by writing the trace of $\Theta^{\mu\nu}$ in d dimensions:

$$\Theta^\mu{}_\mu = -\frac{4-d}{4}(F_{\lambda\sigma})^2 + (1-d)\bar\psi i\not{D}\psi. \tag{19.159}$$

The one-loop matrix element of this quantity proportional to two powers of the background field arises from the three diagrams shown in Fig. 19.10. Since the second term on the right-hand side of (19.159) vanishes by the equation of motion of $\psi(x)$, one might expect that this term gives zero contribution to the trace. Indeed, it is easy to check that the first two diagrams in Fig.

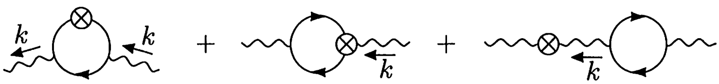
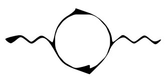


Figure 19.10. One-loop diagrams contributing to the anomalous trace of $\Theta^{\mu\nu}$.

19.10 cancel: These diagrams have the same structure, since the first has an extra propagator and an extra factor $\not{p}$ from the operator matrix element, and opposite overall signs.

The first term on the left-hand side of (19.159) is unexpected, since it apparently vanishes in four dimensions. However, the fermion loop diagram is divergent, and in dimensional regularization, this introduces a factor $1/(2-d/2)$. As a result, this diagram gives a nonzero contribution to the operator matrix element. In massless QED, the fermion loop diagram has the value



$$= -i(k^2 g^{\mu\nu} - k^\mu k^\nu)\frac{4}{3(4\pi)^2}\big(\Gamma(2-\tfrac{d}{2}) + (\text{finite})\big). \quad (19.160)$$

Then the complete expression for the third diagram in Fig. 19.10 is

$$\int \frac{d^4k}{(2\pi)^4} A_\mu(-k)\Big(-2\frac{4-d}{4}\Big)(k^2 g^{\mu\nu} - k^\mu k^\nu)\frac{-i}{k^2}\Big(-ik^2\frac{4}{3(4\pi)^2}\frac{1}{2-d/2}\Big)A_\nu(k). \quad (19.161)$$

This is of the form of (19.158), with

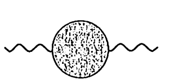
$$C = \frac{1}{12\pi^2}, \quad (19.162)$$

which is indeed the first β function coefficient in massless QED.

This discussion generalizes to QCD and other gauge theories. In a non-Abelian gauge theory, $\Theta^{\mu\nu}$ is given by the obvious generalization of (19.150) with the Abelian field-strength tensor $F_{\mu\nu}$ replaced by the non-Abelian expression $F^a_{\mu\nu}$. The trace of $\Theta^{\mu\nu}$ is again given by

$$\Theta^\mu{}_\mu = -\frac{4-d}{4}(F^a_{\lambda\sigma})^2, \quad (19.163)$$

plus terms that vanish by the equations of motion. In the background field gauge, the one-loop diagrams with the operator $\Theta^\mu{}_\mu$ inserted into the loop cancel as above. We saw in Section 16.6 that the two-point functions in this gauge sum to



$$= -i(k^2 g^{\mu\nu} - k^\mu k^\nu)\Big[\frac{-b_0}{(4\pi)^2}\Big]\big(\Gamma(2-\tfrac{d}{2}) + (\text{finite})\big), \quad (19.164)$$

where $\beta(g) = -b_0 g^3/(4\pi)^2$. Following through the logic of the previous paragraph, we again find the result (19.158) with the identification of C as the first β function coefficient.

As with the axial vector anomaly, the trace anomaly can be found in many different ways. For each possible method of regulating a quantum field theory, there is a derivation of the trace anomaly that exploits the possible pathology of that particular regulator. For example, if one uses a Pauli-Villars regulator with heavy fermions to cancel the divergence of the QED fermion loop diagram, the heavy fermions Ψ contribute a term $M\overline{\Psi}\Psi$ to the trace of $\Theta^{\mu\nu}$. The loop diagram with this term inserted turns out to have a finite limit as $M \to \infty$, which precisely reproduces the trace anomaly. This computation is worked out in Problem 19.4.

As with the axial vector anomaly, each derivation of the anomaly with a different regulator, taken individually, seems artificial, as if there were a problem with the field theory that we are not quite clever enough to fix. Eventually, though, we are forced to conclude that the quantum field theory is trying to tell us something. The anomalous symmetries of the classical theory cannot be promoted to symmetries of the quantum theory. Instead, the anomalous conservation laws require profound and qualitative changes in the theory from the classical to the quantum level.

Problems

19.1 Fermion number nonconservation in parallel E and B fields.

(a) Show that the Adler-Bell-Jackiw anomaly equation leads to the following law for global fermion number conservation: If N_R and N_L are, respectively, the numbers of right- and left-handed massless fermions, then

$$\Delta N_R - \Delta N_L = -\frac{e^2}{2\pi^2}\int d^4x\, \mathbf{E}\cdot\mathbf{B}.$$

To set up a solvable problem, take the background field to be $A^\mu = (0, 0, Bx^1, A)$, with B constant and A constant in space and varying only adiabatically in time.

(b) Show that the Hamiltonian for massless fermions represented in the components (3.36) is

$$H = \int d^3x \Big[\psi_R^\dagger(-i\boldsymbol{\sigma}\cdot\mathbf{D})\psi_R - \psi_L^\dagger(-i\boldsymbol{\sigma}\cdot\mathbf{D})\psi_L\Big],$$

with $D^i = \nabla^i - ieA^i$. Concentrate on the term in the Hamiltonian that involves right-handed fermions. To diagonalize this term, one must solve the eigenvalue equation $-i\boldsymbol{\sigma}\cdot\mathbf{D}\psi_R = E\psi_R$.

(c) The ψ_R eigenvectors can be written in the form

$$\psi_R = \begin{pmatrix}\phi_1(x^1)\\ \phi_2(x^1)\end{pmatrix} e^{i(k_2x^2+k_3x^3)}.$$

The functions ϕ_1 and ϕ_2, which depend only on x^1, obey coupled first-order differential equations. Show that, when one of these functions is eliminated, the other obeys the equation of a simple harmonic oscillator. Use this observation to find the single-particle spectrum of the Hamiltonian. Notice that the eigenvalues do not depend on k_2.

(d) If the system of fermions is set up in a box with sides of length L and periodic boundary conditions, the momenta k_2 and k_3 will be quantized:

$$k_i = \frac{2\pi n_i}{L}.$$

By looking back to the harmonic oscillator equation in part (c), show that the condition that the center of the oscillation is inside the box leads to the condition

$$k_2 < eBL.$$

Combining these two conditions, we see that each level found in part (c) has a degeneracy of

$$\frac{eL^2 B}{2\pi}.$$

(e) Now consider the effect of changing the background A adiabatically by an amount (19.37). Show that the vacuum loses right-handed fermions. Repeating this analysis for the left-handed spectrum, one sees that the vacuum gains the same number of left-handed fermions. Show that these numbers are in accord with the global nonconservation law given in part (a).

19.2 Weak decay of the pion.

(a) In the effective Lagrangian for semileptonic weak interactions (18.28), the hadronic part of the operator is a left-handed current involving the u and d quarks. Show that this current is related to the quark currents of Section 19.3 as follows:

$$\bar{u}_L \gamma^\mu d_L = \frac{1}{2}(j^{\mu 1} + i j^{\mu 2} - j^{\mu 51} - i j^{\mu 52}),$$

where 1, 2 are isospin indices. Using this identification and (19.88), show that the amplitude for the decay $\pi^+ \to \ell^+ \nu$ is given by

$$i\mathcal{M} = G_F f_\pi \bar{u}(q) \not{p}(1 - \gamma^5) v(k),$$

where p, k, q are the momenta of the π^+, ℓ^+, ν.

(b) Compute the decay rate of the pion. Show that this rate vanishes in the limit of zero lepton mass, and that the relative rate of pion decay to muons and electrons is given by

$$\frac{\Gamma(\pi^+ \to e^+\nu)}{\Gamma(\pi^+ \to \mu^+\nu)} = \left(\frac{m_e}{m_\mu}\right)^2 \frac{(1 - m_e^2/m_\pi^2)^2}{(1 - m_\mu^2/m_\pi^2)^2} = 10^{-4}.$$

From the measured pion lifetime, $\tau_\pi = 2.6 \times 10^{-8}$ sec, and the pion and muon masses, $m_\pi = 140$ MeV, $m_\mu = 106$ MeV, determine the value of f_π.

19.3 Computation of anomaly coefficients.

(a) Consider a product $r_1 \times r_2$ of $SU(n)$ representations, which is decomposed into irreducible representations as in (15.98). Using the explicit form of the generators given in (15.99), show that the anomaly coefficients satisfy

$$d_1 A(r_2) + d_2 A(r_1) = \sum_i A(r_i).$$

(b) As we saw in Problem 15.5, the two-index symmetric and antisymmetric tensors form irreducible representations of $SU(n)$, which we will call s and a, respectively. In $SU(3)$, the representation a is three-dimensional. Show that it is equivalent to the $\bar{3}$. Compute the anomaly coefficients for a and s, making use of the identity in part (a).

(c) Since $SU(n)$ has a unique three-index symmetric tensor d^{abc} which is already nonvanishing in an $SU(3)$ subgroup, we can compute the anomaly coefficient in $SU(n)$ by restricting our attention to three generators in this subgroup. By decomposing $SU(n)$ representations into $SU(3)$ representations, compute the anomaly coefficients for a and s in $SU(n)$ and derive Eq. (19.141). Find the anomaly coefficient of the j-index totally antisymmetric tensor representation of $SU(n)$. Why does the result always vanish when $2j = n$?

19.4 Large fermion mass limits. In the text, we derived the Adler-Bell-Jackiw and trace anomalies by the use of dimensional regularization. As an alternative, one could imagine deriving these results using Pauli-Villars regularization. In that technique, one regularizes the value of a fermion loop integral by subtracting the value of the same loop diagram computed with fermions Ψ of large mass M. The parameter M plays the role of the cutoff and should be taken to infinity at the end of the calculation. The anomalies arise because some pieces of the diagrams computed for very heavy fermions do not disappear in the limit $M \to \infty$. These nontrivial $M \to \infty$ limits are interesting in their own right and can have physical applications (for example, in part (c) of the Final Project for Part III).

(a) Show that the Adler-Bell-Jackiw anomaly equation is equivalent to the following large-mass limit of a fermion matrix element between the vacuum and a two-photon state:

$$\lim_{M\to\infty} \left\{ \langle p, k|\, 2iM\bar{\Psi}\gamma^5\Psi\, |0\rangle \right\} = -\frac{e^2}{2\pi^2}\epsilon^{\alpha\nu\beta\lambda}p_\alpha\epsilon^*_\nu(p)k_\beta\epsilon^*_\lambda(k).$$

(b) Show that the trace anomaly, at one-loop order, is equivalent to the following large-mass limit:

$$\lim_{M\to\infty} \left\{ \langle p, k|\, M\bar{\Psi}\Psi\, |0\rangle \right\} = +\frac{e^2}{6\pi^2}\left[p\cdot k\,\epsilon^*(p)\cdot\epsilon^*(k) - p\cdot\epsilon^*(k)\,k\cdot\epsilon^*(p)\right].$$

(c) Show that the matrix element in part (a) is ultraviolet-finite before the $M \to \infty$ limit is taken. Evaluate the matrix element explicitly at one-loop order and verify the limit claimed in part (a).

(d) The matrix element in part (b) has a potential ultraviolet divergence. However, show that the coefficient of $(p\cdot\epsilon^*(k)k\cdot\epsilon^*(p))$ is ultraviolet-finite, and that the rest of the expression is determined by gauge invariance. Compute the full matrix element using dimensional regularization as a gauge-invariant regulator and verify the result claimed in part (b).

Chapter 20

Gauge Theories with Spontaneous Symmetry Breaking

In the course of this book, we have discussed three distinct fashions in which symmetries can be realized in a quantum field theory. The simplest case is a global symmetry that is manifest, leading to particle multiplets with restricted interactions. A second possibility is a global symmetry that is spontaneously broken. Then, as discussed in Chapter 11,* the symmetry currents are still conserved and interactions are similarly restricted, but the vacuum state does not respect the symmetry and the particles do not form obvious symmetry multiplets. Instead, such a theory contains massless particles, Goldstone bosons, one for each generator of the spontaneously broken symmetry. The third case is that of a local, or gauge, symmetry. As we saw in Chapter 15, such a symmetry requires the existence of a massless vector field for each symmetry generator, and the interactions among these fields are highly restricted.

It is now only natural to consider a fourth possibility: What happens if we include both local gauge invariance and spontaneous symmetry breaking in the same theory? In this chapter and the next, we will find that this combination of ingredients leads to new possibilities for the construction of quantum field theory models. We will see that spontaneous symmetry breaking requires gauge vector bosons to acquire mass. However, the interactions of these massive bosons are still constrained by the underlying gauge symmetry, and these constraints can have observable consequences.

In elementary particle physics, the principal application of spontaneously broken local symmetry is in the currently accepted model of weak interactions. This model, due to Glashow, Weinberg, and Salam, is introduced in Section 20.2. There we will see that it makes a number of precise and successful predictions for weak interaction phenomena. Remarkably, this model also unifies the weak interactions with electromagnetism in a single larger gauge theory.

*Section 11.1 is necessary background for the present chapter, but the rest of Chapter 11 is not.

20.1 The Higgs Mechanism

In this section we analyze some simple examples of gauge theories with spontaneous symmetry breaking. We begin with an Abelian gauge theory, and then study several examples of non-Abelian models.

An Abelian Example

As our first example, consider a complex scalar field coupled both to itself and to an electromagnetic field:

$$\mathcal{L} = -\tfrac{1}{4}(F_{\mu\nu})^2 + |D_\mu \phi|^2 - V(\phi), \tag{20.1}$$

with $D_\mu = \partial_\mu + ieA_\mu$. This Lagrangian is invariant under the local $U(1)$ transformation

$$\phi(x) \to e^{i\alpha(x)}\phi(x), \qquad A_\mu(x) \to A_\mu(x) - \frac{1}{e}\partial_\mu \alpha(x). \tag{20.2}$$

If we choose the potential in $\mathcal{L}$ to be of the form

$$V(\phi) = -\mu^2 \phi^* \phi + \frac{\lambda}{2}(\phi^*\phi)^2, \tag{20.3}$$

with $\mu^2 > 0$, the field ϕ will acquire a vacuum expectation value and the $U(1)$ global symmetry will be spontaneously broken. The minimum of this potential occurs at

$$\langle \phi \rangle = \phi_0 = \left(\frac{\mu^2}{\lambda}\right)^{1/2}, \tag{20.4}$$

or at any other value related by the $U(1)$ symmetry (20.2).

Let us expand the Lagrangian (20.1) about the vacuum state (20.4). Decompose the complex field $\phi(x)$ as

$$\phi(x) = \phi_0 + \frac{1}{\sqrt{2}}\big(\phi_1(x) + i\phi_2(x)\big). \tag{20.5}$$

The potential (20.3) is rewritten

$$V(\phi) = -\frac{1}{2\lambda}\mu^4 + \frac{1}{2}\cdot 2\mu^2\phi_1^2 + \mathcal{O}(\phi_i^3), \tag{20.6}$$

so that the field ϕ_1 acquires the mass $m = \sqrt{2}\mu$ and ϕ_2 is the massless Goldstone boson. So far, this whole analysis follows that in Section 11.1.

But now consider how the kinetic energy term of ϕ is transformed. Inserting the expansion (20.5), we rewrite

$$|D_\mu\phi|^2 = \tfrac{1}{2}(\partial_\mu\phi_1)^2 + \tfrac{1}{2}(\partial_\mu\phi_2)^2 + \sqrt{2}e\phi_0 \cdot A_\mu\partial^\mu\phi_2 + e^2\phi_0^2 A_\mu A^\mu + \cdots, \tag{20.7}$$

where we have omitted terms cubic and quartic in the fields A_μ, ϕ_1, and ϕ_2. The last term written explicitly in (20.7) is a photon mass term

$$\Delta\mathcal{L} = \tfrac{1}{2}m_A^2 A_\mu A^\mu, \tag{20.8}$$

where the mass

$$m_A^2 = 2e^2\phi_0^2 \tag{20.9}$$

arises from the nonvanishing vacuum expectation value of ϕ. Notice that the sign of this mass term is correct; the physical spacelike components of A^μ appear in (20.8) as

$$\Delta\mathcal{L} = -\tfrac{1}{2}m_A^2(A^i)^2,$$

with the correct sign for a potential energy term.

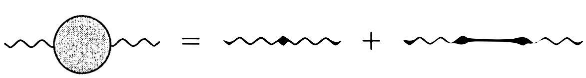
In Chapter 7, and again in Chapter 16 for the non-Abelian case, we argued that a gauge boson cannot obtain a mass, unless this mass term is associated with a pole in the vacuum polarization amplitude. There is a counterexample to this result in two-dimensional spacetime; there, as we saw in Section 19.1, a pole of the required form can arise from the infrared singularity generated by a massless fermion pair. However, in four dimensions, a pole in the vacuum polarization amplitude can be created only by a massless scalar particle. Typically, in situations with unbroken symmetry, no such particle is available.

However, a model with a spontaneously broken continuous symmetry must have massless Goldstone bosons. These scalar particles have the quantum numbers of the symmetry currents, and therefore have just the right quantum numbers to appear as intermediate states in the vacuum polarization. In the model we are now discussing, we can see this pole arise explicitly in the following way: The third term in Eq. (20.7) couples the gauge boson directly to the Goldstone boson ϕ_2; this gives a vertex of the form



$$= i\sqrt{2}e\phi_0(-ik^\mu) = m_A k^\mu. \tag{20.10}$$

If we also treat the mass term (20.8) as a vertex in perturbation theory, then the leading-order contributions to the vacuum polarization amplitude give the expression



$$\begin{aligned} &= im_A^2 g^{\mu\nu} + (m_A k^\mu)\frac{i}{k^2}(-m_A k^\nu) \\ &= im_A^2\Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\Big). \end{aligned} \tag{20.11}$$

The Goldstone boson supplies exactly the right pole to make the vacuum polarization amplitude properly transverse.

Although the Goldstone boson plays an important formal role in this theory, it does not appear as an independent physical particle. The easiest way to see this is to make a particular choice of gauge, called the *unitarity gauge*. Using the local $U(1)$ gauge symmetry (20.2), we can choose $\alpha(x)$ in such a way that $\phi(x)$ becomes real-valued at every point x. With this choice, the field ϕ_2 is removed from the theory. The Lagrangian (20.1) becomes

$$\mathcal{L} = -\tfrac{1}{4}(F_{\mu\nu})^2 + (\partial_\mu\phi)^2 + e^2\phi^2 A_\mu A^\mu - V(\phi). \tag{20.12}$$

If the potential $V(\phi)$ favors a nonzero vacuum expectation value of ϕ, the gauge field acquires a mass; it also retains a coupling to the remaining, physical field ϕ_1.

This mechanism, by which spontaneous symmetry breaking generates a mass for a gauge boson, was explored and generalized to the non-Abelian case by Higgs, Kibble, Guralnik, Hagen, Brout, and Englert, and is now known as the *Higgs mechanism.* However, this mechanism had an earlier application to the theory of superconductivity. In Chapter 8, we constructed the Landau description of a second-order phase transition. To describe a superconductor, Landau and Ginzburg coupled this theory to an external electromagnetic field; they obtained precisely the Lagrangian (20.1). Since the gauge field acquires a nonzero mass, external electromagnetic fields penetrate a superconductor only to the depth m_A^{-1}. This explains the *Meissner effect*, the observed exclusion of macroscopic magnetic fields from a superconductor.

The role of the Goldstone boson in the Higgs mechanism is intricate, and seems mysterious at this level of the discussion. We first saw that the involvement of the Goldstone boson is necessary, as a matter of principle, in order for the gauge boson to acquire a mass. We then saw that the Goldstone boson can be formally eliminated from the theory. However, we might argue that the Goldstone boson has not completely disappeared. A massless vector boson has only two physical polarization states; we saw in Chapter 16 that the longitudinal polarization state cannot be produced, and appears in the formalism only to cancel other unphysical contributions. However, a massive vector boson must have three physical polarization states: In its rest frame, it is a spin-1 object, which can make no distinction between transverse and longitudinal polarizations. It is tempting to say that the gauge boson acquired its extra degree of freedom by *eating* the Goldstone boson. In Sections 21.1 and 21.2 we will clarify this picture, by studying the quantization and gauge-fixing of spontaneously broken gauge theories.

Systematics of the Higgs Mechanism

The Higgs mechanism extends straightforwardly to systems with non-Abelian gauge symmetry. It is not difficult to derive the general relation by which a set of scalar field vacuum expectation values leads to the appearance of gauge boson masses. Let us work out this relation and then apply it in a number of examples.

Consider a system of scalar fields ϕ_i that appear in a Lagrangian invariant under a symmetry group G, represented by the transformation

$$\phi_i \to (1 + i\alpha^a t^a)_{ij}\phi_j. \tag{20.13}$$

It is convenient to write the ϕ_i as real-valued fields, for example, writing n complex fields as $2n$ real fields. Then the group representation matrices t^a must be pure imaginary and, since they are Hermitian, antisymmetric. Let us

write

$$t^a_{ij} = iT^a_{ij}, \tag{20.14}$$

so that the T^a are real and antisymmetric.

If we promote the symmetry group G to a local gauge symmetry, the covariant derivative on the ϕ_i is

$$D_\mu\phi = (\partial_\mu - igA^a_\mu t^a)\phi = (\partial_\mu + gA^a_\mu T^a)\phi.$$

Then the kinetic energy term for the ϕ_i is

$$\tfrac{1}{2}(D_\mu\phi_i)^2 = \tfrac{1}{2}(\partial_\mu\phi_i)^2 + gA^a_\mu(\partial_\mu\phi_i T^a_{ij}\phi_j) + \tfrac{1}{2}g^2 A^a_\mu A^{b\mu}(T^a\phi)_i(T^b\phi)_i. \tag{20.15}$$

Now let the ϕ_i acquire vacuum expectation values

$$\langle\phi_i\rangle = (\phi_0)_i, \tag{20.16}$$

and expand the ϕ_i about these values. The last term in Eq. (20.15) contains a term with the structure of a gauge boson mass,

$$\Delta\mathcal{L} = \tfrac{1}{2}m^2_{ab}A^a_\mu A^{b\mu}, \tag{20.17}$$

with the mass matrix

$$m^2_{ab} = g^2(T^a\phi_0)_i(T^b\phi_0)_i. \tag{20.18}$$

This matrix is positive semidefinite, since any diagonal element, in any basis, has the form

$$m^2_{aa} = g^2(T^a\phi_0)^2 \geq 0 \qquad \text{(no sum)}.$$

Thus, generically, all of the gauge bosons will receive positive masses. However, it may be that some particular generator T^a of G leaves the vacuum invariant:

$$T^a\phi_0 = 0.$$

In that case, the generator T^a will give no contribution to (20.18), and the corresponding gauge boson will remain massless.

As in the Abelian case, the gauge boson propagator receives a contribution from the Goldstone bosons, which is necessary to make the vacuum polarization amplitude transverse. To compute this contribution, we need the vertex that mixes gauge bosons and Goldstone bosons. This comes from the second term of the Lagrangian (20.15). When we insert the vacuum expectation value of the scalar field (20.16), this term becomes

$$\Delta\mathcal{L} = gA^a_\mu\partial_\mu\phi_i(T^a\phi_0)_i. \tag{20.19}$$

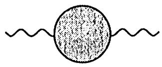
This interaction term does not involve all of the components of ϕ—only those that are parallel to a vector $T^a\phi_0$ for some choice of T^a. These vectors represent the infinitesimal rotations of the vacuum; thus the components ϕ_i that appear in (20.19) are precisely the Goldstone bosons. Using the fact that these

bosons are massless, we can compute the counterpart, for the non-Abelian case, of the Goldstone boson diagram in Eq. (20.11):



$$= \sum_j \big(gk^\mu (T^a\phi_0)_j\big)\frac{i}{k^2}\big(-gk^\nu (T^b\phi_0)_j\big). \qquad (20.20)$$

The sum runs over those components j with a nonzero projection onto the space spanned by the $T^a\phi_0$, or equally well, over all j. This diagram is therefore proportional to the mass matrix (20.18). Combining this expression with the contribution to the vacuum polarization from (20.17), we find a properly transverse result,



$$= i m_{ab}^2 \Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\Big), \qquad (20.21)$$

where m_{ab}^2 is given by Eq. (20.18).

Non-Abelian Examples

Let us now apply this general formalism to some specific examples of non-Abelian gauge theories. Consider first a model with an $SU(2)$ gauge field coupled to a scalar field ϕ that transforms as a spinor of $SU(2)$. The covariant derivative acting on ϕ is

$$D_\mu \phi = (\partial_\mu - igA_\mu^a \tau^a)\phi, \qquad (20.22)$$

where $\tau^a = \sigma^a/2$. The square of this expression is the scalar field kinetic energy term.

If ϕ acquires a vacuum expectation value, we can use the freedom of $SU(2)$ rotations to write this expectation value in the form

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}. \qquad (20.23)$$

Then the gauge boson masses arise from

$$|D_\mu\phi|^2 = \frac{1}{2}g^2 \begin{pmatrix} 0 & v \end{pmatrix} \tau^a\tau^b \begin{pmatrix} 0 \\ v \end{pmatrix} A_\mu^a A^{b\mu} + \cdots. \qquad (20.24)$$

We can symmetrize the matrix product under the interchange of a and b; using $\{\tau^a, \tau^b\} = \frac{1}{2}\delta^{ab}$, we find the mass term

$$\Delta\mathcal{L} = \frac{g^2v^2}{8} A_\mu^a A^{a\mu}. \qquad (20.25)$$

All three gauge bosons receive the mass

$$m_A = \frac{gv}{2}, \qquad (20.26)$$

signaling that all three generators of $SU(2)$ are broken equally well by (20.23).

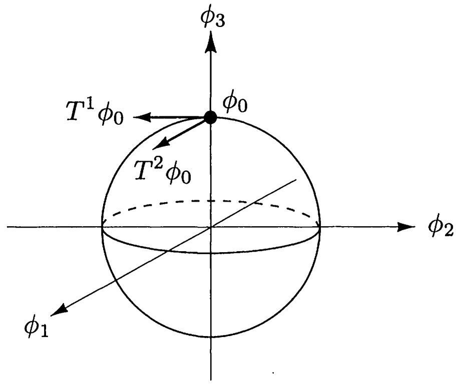


Figure 20.1. The space of configurations for a scalar field in the vector representation of $SU(2)$. When the $SU(2)$ symmetry is spontaneously broken, the allowed vacuum states lie on a spherical surface. If the vacuum expectation value ϕ_0 lies in the 3 direction, then the generator T^3 leaves ϕ_0 unchanged, while T^1 and T^2 rotate ϕ_0 in the directions shown.

What if we had taken ϕ to transform according to the vector representation of $SU(2)$? If we take ϕ to be a real-valued vector under $SU(2)$, we must assign it the covariant derivative

$$(D_\mu\phi)_a = \partial_\mu\phi_a + g\epsilon_{abc}A^b_\mu\phi_c. \tag{20.27}$$

Again, the ϕ kinetic energy term is the square of this object, and so, if ϕ acquires a vacuum expectation value, we find the gauge boson mass term

$$\Delta\mathcal{L} = \frac{1}{2}(D_\mu\phi)^2 = \frac{g^2}{2}\big(\epsilon_{abc}A^b_\mu(\phi_0)_c\big)^2 + \cdots. \tag{20.28}$$

If a vector of $SU(2)$ acquires an expectation value ϕ_0, we can choose our coordinates so that this vector points in any particular direction in the internal space. We will choose it to point in the 3 direction, as indicated in Fig. 20.1:

$$\langle\phi_c\rangle = (\phi_0)_c = V\delta_{c3}. \tag{20.29}$$

Inserting (20.29) into (20.28), we find

$$\Delta\mathcal{L} = \frac{g^2}{2}V^2\big(\epsilon_{ab3}A^b_\mu\big)^2 = \frac{g^2}{2}V^2\big((A^1_\mu)^2 + (A^2_\mu)^2\big). \tag{20.30}$$

The gauge bosons corresponding to the generators 1 and 2 acquire masses

$$m_1 = m_2 = gV, \tag{20.31}$$

while the boson corresponding to the generator 3 remains massless. It is easy to see the reason for this distinction by glancing at Fig. 20.1. The vacuum expectation value of ϕ_c destroys the symmetry of rotation about the axes 1 and 2, but it preserves the symmetry of rotation about the 3 axis. As we saw

in our general analysis, gauge bosons corresponding to unbroken symmetry generators remain massless.

It is interesting that this model contains both massive and massless gauge bosons, with the distinction between these bosons created by spontaneous symmetry breaking. If we interpret the massive bosons as W bosons and the massless gauge boson as the photon, it is tempting to interpret this theory as a unified model of weak and electromagnetic interactions. Georgi and Glashow proposed this model as a serious candidate for the theory of weak interactions.† However, Nature chooses a different model, which we will discuss in the next section.

We turn next to a more complicated example. Consider an $SU(3)$ gauge theory with a scalar field in the adjoint representation. The covariant derivative of ϕ takes the form

$$D_\mu \phi_a = \partial_\mu \phi_a + g f_{abc} A^b_\mu \phi_c, \tag{20.32}$$

and so the gauge field masses arise from the term

$$\Delta\mathcal{L} = \frac{g^2}{2}\left(f_{abc} A^b_\mu \phi_c\right)^2. \tag{20.33}$$

We can write this more clearly by defining the quantity

$$\Phi = \phi_c t^c, \tag{20.34}$$

where t^c are the 3×3 traceless Hermitian matrices that represent the generators of $SU(3)$. Using this notation and the definition (15.68) of the structure constants, we can rewrite the mass term (20.33) as

$$\Delta\mathcal{L} = -g^2 \operatorname{tr}\left[[t^a, \Phi][t^b, \Phi]\right] A^a_\mu A^{b\mu}. \tag{20.35}$$

Now let Φ acquire a vacuum expectation value

$$\langle \Phi \rangle = \Phi_0. \tag{20.36}$$

Since Φ_0 is a traceless Hermitian matrix, we should analyze its effects by diagonalizing it. In principle, Φ_0 could have three arbitrary eigenvalues that sum to zero. However, when one minimizes explicit potential energy functions, one often finds expectation values that preserve some of the original symmetry. We will consider two examples.

First, Φ_0 might have the orientation

$$\Phi_0 = |\phi| \cdot \begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix}. \tag{20.37}$$

†H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **28**, 1494 (1972).

This matrix commutes with the four $SU(3)$ generators

$$t^a = \begin{pmatrix} \tau^a & 0 \\ 0 & 0 \end{pmatrix}, \qquad t^8 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix}. \tag{20.38}$$

Thus, the expectation value (20.37) breaks $SU(3)$ spontaneously to $SU(2) \times U(1)$ and leaves the gauge bosons corresponding to these four generators massless. The remaining four generators of $SU(3)$,

$$\begin{aligned} t^4 &= \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad & t^5 &= \frac{1}{2}\begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\ t^6 &= \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad & t^7 &= \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \end{aligned} \tag{20.39}$$

acquire the masses

$$m^2 = \big(3g|\phi|\big)^2, \tag{20.40}$$

as one can check by substituting these matrices into Eq. (20.35).

Another possible orientation for Φ_0 is

$$\Phi_0 = |\phi| \cdot \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}. \tag{20.41}$$

In this case, only t^3 and t^8 commute with Φ_0, so the original $SU(3)$ symmetry is broken down to $U(1) \times U(1)$. By substituting into (20.35), one can determine that the gauge bosons corresponding to the remaining generators of $SU(3)$ acquire the masses

$$t^1, t^2: \quad m^2 = \big(2g|\phi|\big)^2, \qquad t^4, t^5, t^6, t^7: \quad m^2 = \big(g|\phi|\big)^2. \tag{20.42}$$

Still larger symmetry groups offer a wider variety of symmetry-breaking patterns, and more complex mass matrices. We consider one further example in Problem 20.1.

Formal Description of the Higgs Mechanism

Up to this point, our study of the Higgs mechanism has been based on the explicit analysis of scalar field Lagrangians coupled to gauge fields. Scalar field theories provide the simplest examples of systems with spontaneous symmetry breaking, and the explicit calculations they allow are useful for visualization. But symmetries can be broken in other ways. In the theory of superconductivity, for example, the Abelian gauge invariance of electromagnetism is broken by pairs of electrons that condense in the ground state of a metal. In Section 19.3, we argued that, in the approximation that quark masses are very small, QCD possesses global symmetries that are spontaneously broken by a

condensation of quark-antiquark pairs. In these examples, spontaneous symmetry breaking is the result of strong interactions beyond perturbation theory. We would like to understand whether these more general mechanisms of spontaneous symmetry breaking can also give mass to vector bosons, and, if so, how the masses can be calculated.

To carry out this analysis, we will need to abstract several ideas from the preceding discussion. First, we will discuss in general terms the relations between gauge bosons, Goldstone bosons, and global symmetry currents. Then we will use this information to construct the gauge boson mass matrix without making direct use of the Lagrangian.

Consider, first, an arbitrary quantum field theory $\mathcal{L}_0$ with a global symmetry G. In Section 9.6, we derived the Noether current corresponding to the G symmetry by varying the Lagrangian by a local gauge transformation with infinitesimal parameter $\alpha^a(x)$. Transforming with a constant α^a should leave $\mathcal{L}_0$ unchanged. Then the more general variation of $\mathcal{L}_0$ must take the form

$$\delta\mathcal{L}_0 = (\partial_\mu \alpha^a) J^{\mu a}, \tag{20.43}$$

for some set of vector operators $J^{\mu a}$ built from the fields of $\mathcal{L}_0$. The variational principle then tells us that

$$\partial_\mu J^{\mu a} = 0. \tag{20.44}$$

We can identify the $J^{\mu a}$ as the Noether currents of the global gauge symmetry.

We can now couple this globally symmetric theory to non-Abelian gauge fields, promoting the global symmetry to a local symmetry. To first order in g, the new Lagrangian should take the form

$$\mathcal{L} = \mathcal{L}_0 - gA^a_\mu J^{\mu a} + \mathcal{O}(A^2). \tag{20.45}$$

To check this, note that the transformation (20.43) compensates the variation due to a gauge transformation of A^a_μ, Eq. (15.46), to leading order in g. The terms of order A^2 and higher can in general be arranged to compensate the higher-order terms in the gauge transformation. Thus, matrix elements involving only one insertion of the gauge field can be evaluated using properties of the Noether currents of the original globally symmetric theory. Note in particular that the conservation law for these currents, Eq. (20.44), guarantees that the Ward identities for these matrix elements are satisfied.

If the global symmetry of the theory $\mathcal{L}_0$ is spontaneously broken, this theory will contain Goldstone bosons, which will stand in a special relation to the Noether currents. At long wavelength, the Goldstone bosons become infinitesimal symmetry rotations of the vacuum, $Q^a \left|0\right\rangle$, where Q^a is the global charge associated with $J^{\mu a}$. Thus, the operators $J^{\mu a}$ have the correct quantum numbers to create Goldstone bosons from the vacuum. Let $\left|\pi_k\right\rangle$ denote a Goldstone boson state. In general, there will be a current $J^{\mu a}$ that can create or destroy this boson; we can parametrize the corresponding matrix element as

$$\left\langle 0\right| J^{\mu a}(x) \left|\pi_k(p)\right\rangle = -ip^\mu F^a{}_k e^{-ip\cdot x}, \tag{20.46}$$

where p^μ is the on-shell momentum of the boson and $F^a{}_k$ is a matrix of constants. The elements $F^a{}_k$ vanish when a denotes a generator of an unbroken symmetry. Then the nonvanishing matrix elements of $F^a{}_k$ connect the currents of the spontaneously broken symmetries to their corresponding Goldstone bosons. Since the currents $J^{\mu a}$ are conserved, we find

$$0 = \partial_\mu \langle 0| J^{\mu a}(x) |\pi_k(p)\rangle = -p^2 F^a{}_k e^{-ip\cdot x}, \tag{20.47}$$

which implies that the bosons with nonzero matrix elements (20.46) satisfy $p^2 = 0$ on shell and so are massless. This is another proof of Goldstone's theorem.‡

Since the scalar field theory that we examined earlier in this section should be a special case of this analysis, we should find there an example of the relation given in Eq. (20.46). Comparing Eqs. (20.15) and (20.45), we see that, for the scalar field theory,

$$J^{\mu a} = \partial_\mu \phi_i T^a_{ij} \phi_j, \tag{20.48}$$

which is indeed the Noether current corresponding to the global gauge symmetry. Inserting the vacuum expectation value (20.16), we find

$$J^{\mu a} = \partial_\mu \phi_i (T^a \phi_0)_i, \tag{20.49}$$

which leads to the set of matrix elements

$$\langle 0| J^{\mu a}(x) |\phi_i(p)\rangle = -ip^\mu (T^a \phi_0)_i e^{-ip\cdot x}. \tag{20.50}$$

Using this relation, we can identify

$$F^a{}_i = T^a_{ij} \phi_{0j} \tag{20.51}$$

for the Higgs mechanism in a weakly coupled scalar field theory. To be more precise, the index i runs over all components of the scalar field ϕ. However, we saw in the discussion below Eq. (20.19) that (20.51) is nonzero only for components ϕ_i that are Goldstone bosons, and only for symmetry generators a that are spontaneously broken. Thus the nonzero components of (20.51) form precisely the structure (20.46).

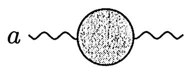
As a concrete illustration of the way that the objects $T^a\phi_0$ link spontaneously broken generators and Goldstone bosons, consider the situation of $SU(2)$ symmetry broken by a scalar field in the vector representation, as in Eq. (20.29) and Fig. 20.1. According to the figure, rotations about the $\hat{1}$ axis tip the vacuum expectation value of ϕ into the $\hat{2}$ direction, rotations about the $\hat{2}$ axis tip this expectation value into the $\hat{1}$ direction, and rotations about $\hat{3}$ leave $\langle\phi\rangle$ invariant. Thus the gauge generators T^1 and T^2 are spontaneously broken, and the scalar field components ϕ^2 and ϕ^1 are the corresponding Goldstone bosons. This accords with the result of computing the elements of $(T^a\phi_0)_i$ explicitly: Using $(T^a)_{bc} = \epsilon^{bac}$, we find

$$(T^a\phi_0)_b = \epsilon_{bac} \langle\phi^c\rangle = V\epsilon^{ba3}. \tag{20.52}$$

‡A special case of this argument appeared in the discussion of Eq. (19.88).

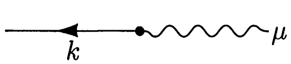
Inserting this result into formula (20.50), we see that the current of each spontaneously broken symmetry creates and destroys its own Goldstone boson.

Now we can use this formalism to study the working of the Higgs mechanism in this general context. Consider the original theory $\mathcal{L}_0$ coupled to gauge bosons of G. To see how the Higgs mechanism operates, we must compute the vacuum polarization amplitude. This amplitude is required by the Ward identity to be transverse, so it is necessarily of the form



$$= i\Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\Big) \cdot (m_{ab}^2 + \mathcal{O}(k^2)). \tag{20.53}$$

It is not easy to compute the nonsingular terms in (20.53) in this general situation, but it is straightforward to compute the singular term, which comes from contributions with an intermediate Goldstone boson. Combining Eqs. (20.45) and (20.46), we see that the amplitude for a gauge boson to convert to a Goldstone boson is



$$= -gk^\mu F^a{}_j. \tag{20.54}$$

Then the pole contribution to the vacuum polarization is

$$= (gk^\mu F^a{}_j)\frac{i}{k^2}(-gk^\nu F^b{}_j). \tag{20.55}$$

Comparing (20.55) with (20.21), we identify

$$m_{ab}^2 = g^2 F^a{}_j F^b{}_j. \tag{20.56}$$

Notice that, in the case in which the symmetry is broken by a scalar field, this result reverts to (20.18). However, Eq. (20.56) applies to any theory of spontaneously broken symmetry, whether the symmetry breaking is apparent from the Lagrangian or whether it requires strong interactions or other nonperturbative effects. It is a general result, then, that any gauge boson coupled to the current of a spontaneously broken symmetry acquires a mass.

20.2 The Glashow-Weinberg-Salam Theory of Weak Interactions

We are now ready to write down the spontaneously broken gauge theory that gives the experimentally correct description of the weak interactions, a model introduced by Glashow, Weinberg, and Salam (GWS). Like the second $SU(2)$ model considered in the previous section, this model gives a unified description of weak and electromagnetic interactions, in which the massless photon corresponds to a particular combination of symmetry generators that remains unbroken.

Again we begin with a theory with $SU(2)$ gauge symmetry. To break the symmetry spontaneously, we introduce a scalar field in the spinor representation of $SU(2)$, as in Eq. (20.22). However, we know that this theory leads

to a system with no massless gauge bosons. We therefore introduce an additional $U(1)$ gauge symmetry. We assign the scalar field a charge $+1/2$ under this $U(1)$ symmetry, so that its complete gauge transformation is

$$\phi \to e^{i\alpha^a \tau^a} e^{i\beta/2}\phi. \tag{20.57}$$

(Here $\tau^a = \sigma^a/2$.) If the field ϕ acquires a vacuum expectation value of the form

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}, \tag{20.58}$$

then a gauge transformation with

$$\alpha^1 = \alpha^2 = 0, \qquad \alpha^3 = \beta \tag{20.59}$$

leaves $\langle\phi\rangle$ invariant. Thus, the theory will contain one massless gauge boson, corresponding to this particular combination of generators. The remaining three gauge bosons will acquire masses from the Higgs mechanism.

Gauge Boson Masses

It is straightforward to work out the details of the mass spectrum by using the methods of the previous section. The covariant derivative of ϕ is

$$D_\mu\phi = \big(\partial_\mu - igA^a_\mu\tau^a - i\tfrac{1}{2}g'B_\mu\big)\phi, \tag{20.60}$$

where A^a_μ and B_μ are, respectively, the $SU(2)$ and $U(1)$ gauge bosons. Since the $SU(2)$ and $U(1)$ factors of the gauge group commute with one another, they can have different coupling constants, which we have called g and g'.

The gauge boson mass terms come from the square of Eq. (20.60), evaluated at the scalar field vacuum expectation value (20.58). The relevant terms are

$$\Delta\mathcal{L} = \frac{1}{2}\begin{pmatrix}0 & v\end{pmatrix}\Big(gA^a_\mu\tau^a + \frac{1}{2}g'B_\mu\Big)\Big(gA^{b\mu}\tau^b + \frac{1}{2}g'B^\mu\Big)\begin{pmatrix}0\\ v\end{pmatrix}. \tag{20.61}$$

If we evaluate the matrix product explicitly, using $\tau^a = \sigma^a/2$, we find

$$\Delta\mathcal{L} = \frac{1}{2}\frac{v^2}{4}\big[g^2(A^1_\mu)^2 + g^2(A^2_\mu)^2 + (-gA^3_\mu + g'B_\mu)^2\big]. \tag{20.62}$$

There are three massive vector bosons, which we will notate as follows:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}(A^1_\mu \mp iA^2_\mu) \qquad \text{with mass} \quad m_W = g\frac{v}{2}; \\ Z^0_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(gA^3_\mu - g'B_\mu) \qquad \text{with mass} \quad m_Z = \sqrt{g^2+g'^2}\,\frac{v}{2}. \end{aligned} \tag{20.63}$$

The fourth vector field, orthogonal to Z^0_μ, remains massless:

$$A_\mu = \frac{1}{\sqrt{g^2+g'^2}}(g'A^3_\mu + gB_\mu) \qquad \text{with mass} \quad m_A = 0. \tag{20.64}$$

We will identify this field with the electromagnetic vector potential.

From now on it will be more convenient to write all expressions in terms of these mass eigenstate fields. Consider, for instance, the coupling of the vector fields to fermions. For a fermion field belonging to a general $SU(2)$ representation, with $U(1)$ charge Y, the covariant derivative takes the form

$$D_\mu = \partial_\mu - igA^a_\mu T^a - ig'YB_\mu. \tag{20.65}$$

In terms of the mass eigenstate fields, this becomes

$$\begin{aligned} D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}(W^+_\mu T^+ + W^-_\mu T^-) - i\frac{1}{\sqrt{g^2+g'^2}}Z_\mu(g^2T^3 - g'^2Y) \\ - i\frac{gg'}{\sqrt{g^2+g'^2}}A_\mu(T^3+Y), \end{aligned} \tag{20.66}$$

where $T^\pm = (T^1 \pm iT^2)$. The normalization is chosen so that, in the spinor representation of $SU(2)$,

$$T^\pm = \tfrac{1}{2}(\sigma^1 \pm i\sigma^2) = \sigma^\pm. \tag{20.67}$$

The last term of Eq. (20.66) makes explicit the fact that the massless gauge boson A_μ couples to the gauge generator (T^3+Y), which generates precisely the symmetry operation (20.59).

To put expression (20.66) into a more useful form, we should identify the coefficient of the electromagnetic interaction as the electron charge e,

$$e = \frac{gg'}{\sqrt{g^2+g'^2}}, \tag{20.68}$$

and identify the electric charge quantum number as

$$Q = T^3 + Y. \tag{20.69}$$

These substitutions, with $Q = -1$ for the electron, give the conventional form of the coupling of the electromagnetic field.

To simplify expression (20.66) further, we define the *weak mixing angle*, θ_w, to be the angle that appears in the change of basis from (A^3, B) to (Z^0, A):

$$\begin{pmatrix} Z^0 \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} A^3 \\ B \end{pmatrix},$$

that is,

$$\cos\theta_w = \frac{g}{\sqrt{g^2+g'^2}}, \qquad \sin\theta_w = \frac{g'}{\sqrt{g^2+g'^2}}. \tag{20.70}$$

Then, with the manipulation in the Z^0 coupling

$$g^2T^3 - g'^2Y = (g^2+g'^2)T^3 - g'^2Q,$$

we can rewrite the covariant derivative (20.66) in the form

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}\left(W_\mu^+ T^+ + W_\mu^- T^-\right) - i\frac{g}{\cos\theta_w} Z_\mu \left(T^3 - \sin^2\theta_w Q\right) - ieA_\mu Q, \tag{20.71}$$

where

$$g = \frac{e}{\sin\theta_w}. \tag{20.72}$$

We see here that the couplings of all of the weak bosons are described by two parameters: the well-measured electron charge e, and a new parameter θ_w. The couplings induced by W and Z exchange will also involve the masses of these particles. However, these masses are not independent, since it follows from Eqs. (20.63) that

$$m_W = m_Z \cos\theta_w. \tag{20.73}$$

All effects of W and Z exchange processes, at least at tree level, can be written in terms of the three basic parameters e, θ_w, and m_W.

Coupling to Fermions

The covariant derivative (20.71) uniquely determines the coupling of the W and Z^0 fields to fermions, once the quantum numbers of the fermion fields are specified. To determine these quantum numbers, we must take account of the fact, mentioned in Section 17.3, that the W boson couples only to left-handed helicity states of quarks and leptons.

At the level of the classical Lagrangian, there is no difficulty in constructing theories in which the left- and right-handed components of a fermion field couple differently to gauge bosons.* Already in Section 3.2 we saw that the kinetic energy term for Dirac fermions splits into separate pieces for the left- and right-handed fields:

$$\bar{\psi} i\partial\!\!\!/ \psi = \bar{\psi}_L i\partial\!\!\!/ \psi_L + \bar{\psi}_R i\partial\!\!\!/ \psi_R. \tag{20.74}$$

When we couple ψ to a gauge field, we can assign ψ_L and ψ_R to different representations of the gauge group. Then the two terms on the right-hand side of (20.74) will contain two different covariant derivatives, and these will imply two different sets of couplings.

In the GWS model, we can use this technique to insure that only the left-handed components of the quark and lepton fields couple to the W bosons. We assign the left-handed fermion fields to doublets of $SU(2)$, while making the right-handed fermion fields singlets under this group. Once we have specified the T^3 value for each fermion field, the value of Y that we must assign follows from Eq. (20.69). This means that the Y assignments will also be different for

*In Section 19.4, we argued that there is a possible problem with this strategy at the level of quantum corrections. We will check below whether the specific model we construct avoids this problem.

the left- and right-handed components of quarks and leptons. For the right-handed fields, $T^3 = 0$, and so we reproduce the standard electric charges by assigning Y to equal the electric charge. For example, for the right-handed u quark field, $Y = +2/3$; for e^-_R, $Y = -1$. For the left-handed fields,

$$E_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \qquad Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \tag{20.75}$$

the assignments $Y = -1/2$ and $Y = +1/6$, respectively, combine with $T^3 = \pm 1/2$ to give the correct electric charge assignments. Since the left- and right-handed fermions live in different representations of the fundamental gauge group, it is often useful to think of these components as distinct particles, which are mixed by the fermion mass terms.

In fact, the construction of fermion mass terms is a serious problem, because all possible such terms are forbidden by global gauge invariances. For example, we cannot write an electron mass term

$$\Delta\mathcal{L} = -m_e\big(\bar{e}_L e_R + \bar{e}_R e_L\big), \tag{20.76}$$

because the fields e_L and e_R belong to different $SU(2)$ representations and have different $U(1)$ charges. For the next few pages, we will ignore this problem by treating all fermion fields as massless. This description will suffice to analyze the structure of the weak interactions at high energies, where the quark and lepton masses can be ignored. At the end of this section we will return to the problem of writing quark and lepton mass terms in the GWS theory. The solution to this problem will reinforce the idea that the left- and right-handed fermion fields are fundamentally independent entities, mixed to form massive fermions by some subsidiary process.

If we ignore fermion masses, the Lagrangian for the weak interactions of quarks and leptons follows directly from the charge assignments given above. The fermion kinetic energy terms for e, ν, u, and d are

$$\mathcal{L} = \overline{E}_L(i\not{D})E_L + \bar{e}_R(i\not{D})e_R + \overline{Q}_L(i\not{D})Q_L + \bar{u}_R(i\not{D})u_R + \bar{d}_R(i\not{D})d_R. \tag{20.77}$$

In each term, the covariant derivative is given by Eq. (20.65), with T^a and Y evaluated in the particular representation to which that fermion field belongs. For example,

$$\overline{Q}_L(i\not{D})Q_L = \overline{Q}_L i\gamma^\mu\big(\partial_\mu - igA^a_\mu\tau^a - i\tfrac{1}{6}g'B_\mu\big)Q_L. \tag{20.78}$$

A right-handed neutrino would have zero coupling both to $SU(2)$ and to $U(1)$, so we have simply omitted this field from Eq. (20.77).

To work out the physical consequences of the fermion-vector boson couplings, we should write Eq. (20.77) in terms of the vector boson mass eigenstates, using the form of the covariant derivative given in Eq. (20.71). Equation (20.77) then takes the form

$$\begin{aligned}\mathcal{L} = {} & \overline{E}_L(i\not{\partial})E_L + \bar{e}_R(i\not{\partial})e_R + \overline{Q}_L(i\not{\partial})Q_L + \bar{u}_R(i\not{\partial})u_R + \bar{d}_R(i\not{\partial})d_R \\ & + g\big(W^+_\mu J^{\mu+}_W + W^-_\mu J^{\mu-}_W + Z^0_\mu J^\mu_Z\big) + eA_\mu J^\mu_{EM},\end{aligned} \tag{20.79}$$

where

$$\begin{aligned}
J_W^{\mu+} &= \frac{1}{\sqrt{2}}\big(\bar{\nu}_L\gamma^\mu e_L + \bar{u}_L\gamma^\mu d_L\big);\\
J_W^{\mu-} &= \frac{1}{\sqrt{2}}\big(\bar{e}_L\gamma^\mu \nu_L + \bar{d}_L\gamma^\mu u_L\big);\\
J_Z^\mu &= \frac{1}{\cos\theta_w}\Big[\bar{\nu}_L\gamma^\mu\big(\tfrac{1}{2}\big)\nu_L + \bar{e}_L\gamma^\mu\big(-\tfrac{1}{2}+\sin^2\theta_w\big)e_L + \bar{e}_R\gamma^\mu\big(\sin^2\theta_w\big)e_R\\
&\qquad + \bar{u}_L\gamma^\mu\big(\tfrac{1}{2}-\tfrac{2}{3}\sin^2\theta_w\big)u_L + \bar{u}_R\gamma^\mu\big(-\tfrac{2}{3}\sin^2\theta_w\big)u_R\\
&\qquad + \bar{d}_L\gamma^\mu\big(-\tfrac{1}{2}+\tfrac{1}{3}\sin^2\theta_w\big)d_L + \bar{d}_R\gamma^\mu\big(\tfrac{1}{3}\sin^2\theta_w\big)d_R\Big];\\
J_{EM}^\mu &= \bar{e}\gamma^\mu(-1)e + \bar{u}\gamma^\mu\big(+\tfrac{2}{3}\big)u + \bar{d}\gamma^\mu\big(-\tfrac{1}{3}\big)d.
\end{aligned}\tag{20.80}$$

Here we have used Eq. (20.67) to simplify the W boson currents. Notice that the current J_{EM}^μ associated with the photon field is indeed the standard electromagnetic current.

Anomaly Cancellation

As we have just seen, there is no difficulty in writing a Lagrangian that couples the GWS gauge bosons to fermions in a chiral fashion. However, these chiral couplings do present a potential problem that appears at the level of one-loop corrections. In Section 19.2, we saw that an axial current that is conserved at the level of the classical equations of motion can acquire a nonzero divergence through one-loop diagrams that couple this current to a pair of gauge bosons. The Feynman diagram that contains this anomalous contribution is a triangle diagram with the axial current and the two gauge currents at its vertices. In a gauge theory in which gauge bosons couple to a chiral current, the dangerous triangle diagrams appear in the one-loop corrections to the three-gauge-boson vertex function. The anomalous terms violate the Ward identity for this amplitude. Thus, as we argued in Section 19.4, theories in which gauge bosons couple to chiral currents can be gauge invariant only if the anomalous contribution somehow disappears. Fortunately, as we saw there, the anomalous terms can be arranged to cancel when one sums over all possible fermion species that can circulate in these diagrams.†

Within the GWS theory, the requirement from experiment that the weak interaction currents are left-handed forced us to choose a chiral gauge coupling. Now we must check that the anomalous terms from the triangle diagrams cancel as required. We will find that they do, but only through a subtle and rather magical interplay of the quantum numbers of quarks and leptons.

The anomalous term of a triangle diagram of three gauge bosons A_μ^a, A_ν^b,

†If you have not read Chapter 19, but you are willing to assume that the fermion triangle diagram contains a contribution that violates gauge invariance, you should still be able to follow the argument that follows.

and A^c_λ is proportional to the group theoretic invariant

$$\mathrm{tr}\big[\gamma^5 t^a \{t^b, t^c\}\big], \tag{20.81}$$

where the trace is taken over all fermion species. The anticommutator comes from taking the sum of two triangle diagrams in which the fermions circle in opposite directions. The factor γ^5 registers the fact that the anomaly is associated with chiral currents; this factor equals -1 for left-handed fermions and $+1$ for right-handed fermions. In theories such as QED or QCD in which the gauge bosons couple equally to right- and left-handed species, the anomalies automatically cancel. This bookkeeping method is a special case of the more general method presented in Section 19.4.

To evaluate the anomalies of the GWS theory, it is easiest to work in the basis of $SU(2) \times U(1)$ gauge bosons, before the mixing to the photon and Z^0 mass eigenstates. It suffices to evaluate the triangle diagrams for massless fermions, so that right- and left-handed fermions have distinct quantum numbers. However, we must consider not only the anomalies of diagrams with three $SU(2) \times U(1)$ gauge bosons, but also diagrams with both weak-interaction gauge bosons and color $SU(3)$ gauge bosons of QCD. If we consider effects of gravity on the weak-interaction gauge theory, there is also a possibly anomalous diagram with one weak-interaction gauge boson and two gravitons. We can omit diagrams, such as the anomaly of three $SU(3)$ bosons or of one $SU(3)$ boson and two gravitons, in which all of the couplings are left-right symmetric. Then the full set of diagrams with possible anomalous terms is shown in Fig. 20.2. All of the possible anomalies must cancel if the Ward identities of the $SU(2) \times U(1)$ gauge theory are to be satisfied.

It is a special property of $SU(2)$ gauge theory that the anomaly of three $SU(2)$ gauge bosons always vanishes; this result follows from the property of Pauli sigma matrices $\{\sigma^a, \sigma^b\} = 2\delta^{ab}$, which implies that the trace (20.81) vanishes. The anomalies containing one $SU(3)$ boson or one $SU(2)$ boson are proportional to

$$\mathrm{tr}[t^a] = 0 \qquad \text{or} \qquad \mathrm{tr}[\tau^a] = 0. \tag{20.82}$$

The remaining nontrivial anomalies are those of one $U(1)$ boson with two $SU(2)$ bosons or two $SU(3)$ bosons, the anomaly of three $U(1)$ bosons, and the gravitational anomaly with one $U(1)$ gauge boson.

The anomaly of one $U(1)$ boson with two $SU(3)$ bosons is proportional to the group theory factor

$$\mathrm{tr}[t^a t^b Y] = \tfrac{1}{2}\delta^{ab} \cdot \sum_q Y_q, \tag{20.83}$$

where the sum runs over left-handed quarks and right-handed quarks, with an extra (-1) for the left-handed contributions. Inserting the charge assignments given above for u_L, d_L, u_R, and d_R, we find

$$\sum_q Y_q = -2 \cdot \tfrac{1}{6} + (\tfrac{2}{3}) + (-\tfrac{1}{3}) = 0. \tag{20.84}$$

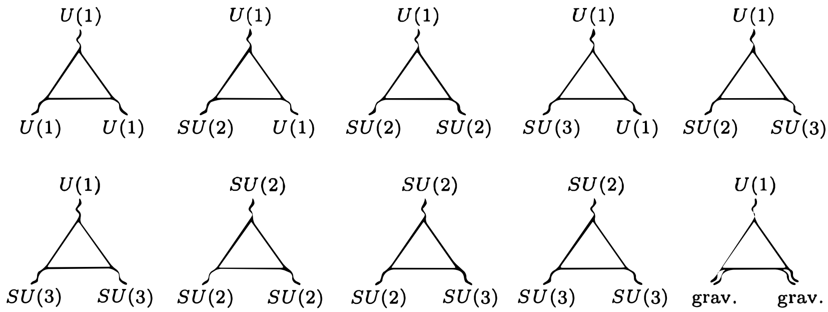


Figure 20.2. Possible gauge anomalies of weak interaction theory. All of these anomalies must vanish for the Glashow-Weinberg-Salam theory to be consistent.

Similarly, the anomaly of a $U(1)$ boson with two $SU(2)$ bosons is proportional to

$$\text{tr}[\tau^a\tau^b Y] = \tfrac{1}{2}\delta^{ab}\sum_{fL} Y_{fL}, \tag{20.85}$$

where the sum runs over the left-handed fermions E_L and Q_L. Thus,

$$\sum_{fL} Y_{fL} = -(-\tfrac{1}{2}) - 3\cdot\tfrac{1}{6} = 0; \tag{20.86}$$

the factor 3 counts the color states of the quarks. The anomaly of three $U(1)$ gauge bosons is proportional to a sum involving left- and right-handed leptons and quarks:

$$\text{tr}[Y^3] = -2(-\tfrac{1}{2})^3 + (-1)^3 - 3\big[2(\tfrac{1}{6})^3 - (\tfrac{2}{3})^3 - (-\tfrac{1}{3})^3\big] = 0. \tag{20.87}$$

Finally, the gravitational anomaly with one $U(1)$ gauge boson is proportional to

$$\text{tr}[Y] = -2(-\tfrac{1}{2}) + (-1) - 3\big[2(\tfrac{1}{6}) - (\tfrac{2}{3}) - (-\tfrac{1}{3})\big] = 0. \tag{20.88}$$

The Glashow-Weinberg-Salam theory is thus a chiral gauge theory that is completely free of axial vector anomalies among the gauge currents. However, the cancellation of anomalies requires that leptons and quarks appear in complete multiplets with the structure of $(E_L, e_R, Q_L, u_R, d_R)$. This set of fields is often called a *generation* of quarks and leptons. The consistency of the theory requires that quarks and leptons appear in Nature in equal numbers, organizing themselves into successive generations in this way.

Experimental Consequences of the GWS Theory

Now that we have a fundamental theory for the coupling of W and Z bosons to fermions, we can work out the consequences of this theory for observable processes mediated by weak bosons. This analysis should reproduce the effective Lagrangian description of the weak interactions used in Chapters 17

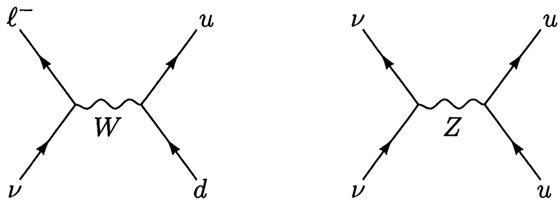


Figure 20.3. Some processes with virtual W and Z boson exchange.

and 18, and also predict additional observable effects of weak boson exchange. In our discussion here, we will derive only the most basic relations in this subject; we do not have space for a systematic survey of the phenomenology of weak interactions. However, we encourage the reader to study the experimental foundations of the weak interactions, which contain many beautiful illustrations of the principles of quantum field theory.‡

At energies low compared to the vector boson masses, the couplings of the weak bosons have their major effects through processes that involve virtual weak boson exchange. These processes are shown in Fig. 20.3. We will derive the Feynman rules for massive gauge bosons in Chapter 21. Meanwhile, it is reasonable to guess that the W and Z boson propagators are given by

$$\big\langle W^{\mu+}(p)W^{\nu-}(-p)\big\rangle = \frac{-ig^{\mu\nu}}{p^2 - m_W^2}, \qquad \langle Z^\mu(p)Z^\nu(-p)\rangle = \frac{-ig^{\mu\nu}}{p^2 - m_Z^2}. \quad (20.89)$$

We will see in Section 21.1 that these propagators give correct expressions for diagrams with W and Z exchange up to terms of order (m_f/m_W), where m_f is a fermion mass.

First consider the W exchange diagram in Fig. 20.3, in the limit of energies low compared to the W mass. We can then neglect the p^2 term in the denominator of the W propagator (20.89). Taking the W coupling from Eq. (20.79), we find that the diagram can be described by the effective Lagrangian

$$\begin{aligned} \Delta\mathcal{L}_W &= \frac{g^2}{m_W^2} J_W^{\mu-} J_{\mu W}^{+} \\ &= \frac{g^2}{2m_W^2}\big(\bar{e}_L\gamma^\mu\nu_L + \bar{d}_L\gamma^\mu u_L\big)\big(\bar{\nu}_L\gamma_\mu e_L + \bar{u}_L\gamma_\mu d_L\big). \end{aligned} \quad (20.90)$$

The coefficient is often written in terms of the *Fermi constant*,

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}. \quad (20.91)$$

The various terms in this effective Lagrangian reproduce the expressions we have already written in Eqs. (17.31), (18.28), and (18.29). Since these interactions among leptons and quarks are mediated by the exchange of an

‡The experimental successes of the theory of weak interactions are reviewed in the book of Commins and Bucksbaum (1983).

electrically charged vector boson, they are called collectively *charged-current* interactions. The effective Lagrangian (20.90) turns out to provide an impressively successful description of the phenomenology of charged-current weak interactions. We have described its use in high-energy neutrino scattering, but it has comparable successes in nuclear β-decay, muon decay, and a variety of other processes.

In a similar way, we can work out the effective Lagrangian resulting from virtual Z^0 exchange. We find

$$\begin{aligned}\Delta\mathcal{L}_Z &= \frac{g^2}{2m_Z^2} J_Z^\mu J_{\mu Z} \\ &= \frac{4G_F}{\sqrt{2}}\Big(\sum_f \bar{f}\gamma^\mu (T^3 - \sin^2\theta_w Q) f\Big)^2,\end{aligned} \tag{20.92}$$

where the sum in the second line runs over all left-handed and right-handed flavors, and we have used relation (20.73) to simplify the prefactor. We say that the effective Lagrangian (20.92) mediates *neutral current* weak interaction processes. Notice that, if we define $SU(2)$ gauge currents as

$$J^{\mu a} = \sum_f \bar{f}\gamma^\mu T^a f, \tag{20.93}$$

then the effective Lagrangians of W and Z exchange can be written together in the simple form

$$\Delta\mathcal{L}_W + \Delta\mathcal{L}_Z = \frac{4G_F}{\sqrt{2}}\Big[(J^{\mu 1})^2 + (J^{\mu 2})^2 + (J^{\mu 3} - \sin^2\theta_w J^\mu_{EM})^2\Big]. \tag{20.94}$$

This expression becomes manifestly invariant under an unbroken global $SU(2)$ symmetry in the limit $g' \to 0$ or $\sin^2\theta_w \to 0$. We will discuss this observation further at the end of this section.

The neutral current effective Lagrangian (20.92) contains terms that couple together all of the various species of quarks and leptons. These terms violate parity, and so distinguish themselves from the effects of strong and electromagnetic interactions. For example, Eq. (20.92) predicts the existence of neutral current deep inelastic neutrino scattering events, in which a high-energy neutrino shatters a nucleon but does not convert to a final-state muon or electron. This process is analyzed in Problem 20.4. Similarly, the neutral-current interaction predicts the presence of parity-violating effects in electron deep inelastic scattering. It also predicts a parity-violating electron-nucleon interaction that should mix atomic energy levels, and a similar parity-violating nucleon-nucleon interaction. Within the GWS theory, the strengths of all of these various effects are predicted in terms of the Fermi constant and one additional parameter, the value of $\sin^2\theta_w$. Thus, the GWS theory can be tested by observing each of these effects and asking whether a single value of this parameter can account for the strengths of all of these disparate processes.

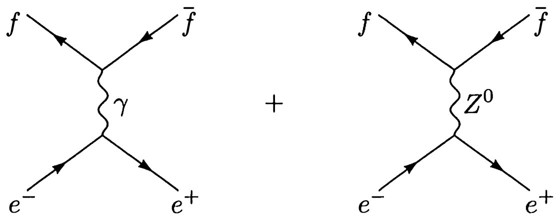


Figure 20.4. Diagrams contributing to the process $e^+e^- \rightarrow f\bar{f}$ in the Glashow-Weinberg-Salam theory.

Further tests of the GWS theory are available at higher energies. The process $e^+e^- \rightarrow f\bar{f}$ is affected in an essential way, since the theory contains a new diagram with s-channel Z^0 exchange, which interferes with the standard photon exchange diagram, as shown in Fig. 20.4. It is straightforward to work out the effects of this interference using the methods of Section 5.2, so we have left this analysis as Problem 20.3.

As the center-of-mass energy approaches m_Z, the Z^0 appears directly as a resonance in the e^+e^- annihilation cross section. Similarly, both the W and the Z can be observed as resonances in quark-antiquark annihilation, viewed as a parton subprocess in proton-antiproton scattering. The positions of these resonances are predicted from G_F, $\sin^2\theta_w$, and the value of e or α, according to Eqs. (20.72) and (20.91). Using these relations, we find

$$m_W^2 = \frac{\pi\alpha}{\sqrt{2}G_F\sin^2\theta_w}, \qquad m_Z^2 = \frac{\pi\alpha}{\sqrt{2}G_F\sin^2\theta_w\cos^2\theta_w}. \tag{20.95}$$

The detailed shape of the Z^0 resonance is shown in Fig. 20.5. The experimental measurements shown are compared to a theoretical curve with the resonance position adjusted for the best fit. The height and width of the resonance are then predicted by the GWS theory. The resonance is broadened to higher energies by processes in which the electron and positron radiate collinear photons before annihilation; this correction was discussed in Problem 5.5.

Because the Lagrangian of the GWS theory treats left- and right-handed fermions as distinct species with completely different quantum numbers, the couplings of the Z^0 to left- and right-handed fermions differ signficantly. One manifestation of this is the presence of a *polarization asymmetry*, a net polarization of fermions produced in the decay $Z^0 \rightarrow f\bar{f}$, or an asymmetry in the inverse process of Z^0 production. This asymmetry can be read directly from the form of the Z^0 current given in (20.80):

$$\begin{aligned} A_{LR}^f &= \frac{\Gamma(Z^0 \rightarrow f_L\bar{f}_R) - \Gamma(Z^0 \rightarrow f_R\bar{f}_L)}{\Gamma(Z^0 \rightarrow f_L\bar{f}_R) + \Gamma(Z^0 \rightarrow f_R\bar{f}_L)} \\ &= \frac{(\frac{1}{2} - |Q_f|\sin^2\theta_w)^2 - (Q_f\sin^2\theta_w)^2}{(\frac{1}{2} - |Q_f|\sin^2\theta_w)^2 + (Q_f\sin^2\theta_w)^2}. \end{aligned} \tag{20.96}$$

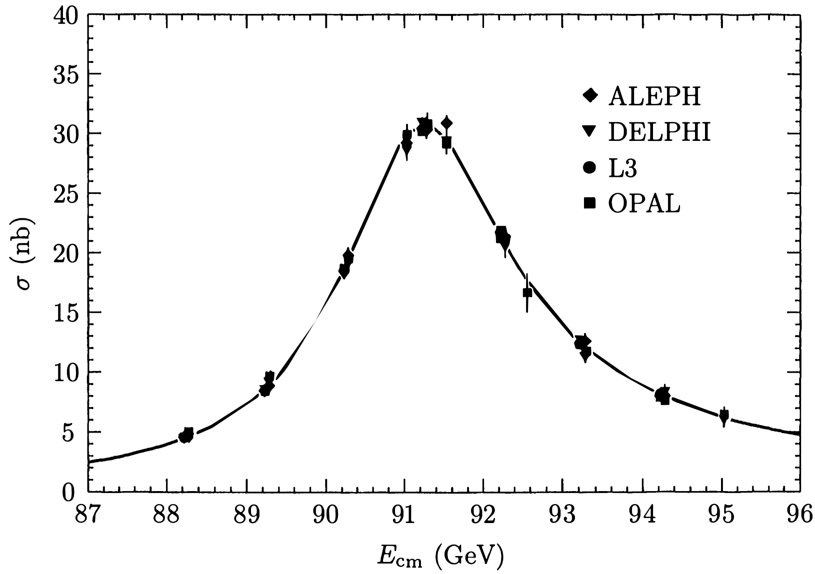


Figure 20.5. The total cross section for e^+e^- annihilation to hadrons for $E_{\rm cm}$ close to the Z^0 boson mass, as measured by the ALEPH, DELPHI, L3, and OPAL experiments and compiled by the Particle Data Group, *Phys. Rev.* **D50**, (1994), Fig. 32.14. References to the original articles are given there. The solid curve is the prediction of the GWS theory.

For a realistic value $\sin^2\theta_w = 0.23$, this expression gives a 15% asymmetry for charged leptons and a 95% asymmetry for d, s, and b quarks. The asymmetry can be checked experimentally for leptons by measuring the polarization of τ leptons at the Z^0 resonance, or by measuring the relative cross sections for producing the resonance using left- versus right-handed electrons. For quarks, the asymmetry can be determined indirectly from the forward-backward production asymmetry on the resonance, as explained in Problem 20.3.

Because the weak neutral current has so many different manifestations, the GWS theory of weak interactions can be subjected to a stringent test by comparing the values of the parameter $\sin^2\theta_w$ needed to account for each of its predicted effects. Table 20.1 presents the values of $\sin^2\theta_w$ extracted from a wide variety of weak interaction neutral current effects and asymmetries. In all cases, one-loop radiative corrections must be included to analyze the experiment at the required level of accuracy. These radiative corrections involve some subtlety. First, one must adopt a specific renormalization convention that defines $\sin^2\theta_w$ and use it consistently in all calculations. The table shows results for two different choices of this convention. In both conventions, the values of weak-interaction observables are taken to be functions of α, G_F, and a third independent parameter. In the first column this parameter is the mass ratio m_W/m_Z, and, following the tree-level expression (20.73), we consider

Table 20.1. Values of $\sin^2\theta_w$ from Weak Interaction Experiments

Observed Quantity or Process	s_W^2	$\sin^2\theta_{w\,\overline{MS}}$
m_Z	0.2247 (21)	0.2320 (6)
m_W	0.2264 (25)	0.2338 (22)
Γ_Z	0.2250(18)	0.2322 (6)
Lepton f-b asymmetries at the Z^0	0.2243 (17)	0.2315 (11)
All pair-production asymmetries at the Z^0	0.2245 (17)	0.2317 (8)
A_{LR}^e at the Z^0	0.2221 (17)	0.2292 (10)
Deep inelastic neutrino scattering	0.2260 (48)	0.233 (5)
Neutrino-proton elastic scattering	0.205 (31)	0.212 (32)
Neutrino-electron elastic scattering	0.224 (9)	0.231 (9)
Atomic parity violation	0.216 (8)	0.223 (8)
Parity violation in inelastic e^- scattering	0.216 (17)	0.223 (18)

The values listed here are obtained by fitting experimental observations by adjusting the value of s_W^2 or $\sin^2\theta_{w\,\overline{MS}}$, taking α and G_F as accurately known parameters. The numbers in parentheses are the standard errors in the last displayed digits. The conversion from the experimentally measured quantities to s_W^2 or $\sin^2\theta_{w\,\overline{MS}}$ depends on the value of the top quark mass and the mass of the Higgs boson. These values assume a top quark mass of 169 GeV and a Higgs mass of 300 GeV; the quoted errors include an uncertainty of 17 GeV in the top quark mass and a range from 60 GeV to 1000 GeV for the Higgs mass. The differences in the relative errors between the two columns reflect the importance of this theoretical uncertainty. Some observables depend weakly on α_s; these values assume $\alpha_s(m_Z) = 0.120 \pm .007$. This table is taken from the article of P. Langacker and J. Erler for the Particle Data Group, *Phys. Rev.* **D50**, 1304 (1994). That article contains a full set of references and a discussion of the sources of uncertainty in these determinations.

this ratio to define a renormalized value of $\sin^2\theta_w$:

$$s_W^2 \equiv 1 - \frac{m_W^2}{m_Z^2}. \tag{20.97}$$

In the second column, the third parameter is $\sin^2\theta_w$ computed from the weak interaction coupling constants defined by minimal subtraction (Eq. (11.77)). The differences between different definitions of $\sin^2\theta_w$ appear at the level of one-loop computations and can reveal interesting physics; this subject is discussed in Section 21.3.

A second subtlety is that the one-loop corrections to weak neutral current processes depend on the value of the t quark mass, which has only recently been determined and is still somewhat poorly known. The dependence on the t quark mass is relatively strong, for interesting reasons that we will discuss in Section 21.3. The one-loop corrections also depend weakly on properties of the particles responsible for the spontaneous symmetry breaking.

We can see from Table 20.1 that a wide variety of effects due to the weak neutral current have been observed, with magnitudes accounted for by a single, consistent value of $\sin^2\theta_w$. This remarkable concordance of theory and experiment gives us confidence that the Glashow-Weinberg-Salam theory is indeed the correct description of weak and electromagnetic interactions.

Fermion Mass Terms

We now return to the problem of writing mass terms for the quarks and leptons. Recall that one cannot put ordinary mass terms into the Lagrangian, because the left- and right-handed components of the various fermion fields have different gauge quantum numbers and so simple mass terms violate gauge invariance. To give masses to the quarks and leptons, we must again invoke the mechanism of spontaneous symmetry breaking.

We began this section by assuming that a scalar field ϕ acquires a vacuum expectation value (20.58), in order to give mass to the W and Z bosons. This scalar field needed to be a spinor under $SU(2)$ and to have $Y = 1/2$ in order to produce the correct pattern of gauge boson masses. With these quantum numbers, we can also write a gauge-invariant coupling linking e_L, e_R, and ϕ, as follows:

$$\Delta\mathcal{L}_e = -\lambda_e \overline{E}_L \cdot \phi\, e_R + \text{h.c.} \tag{20.98}$$

Here the $SU(2)$ indices of the doublets E_L and ϕ are contracted; notice also that the charges Y of the various fields sum to zero. The parameter λ_e is a new dimensionless coupling constant. If we replace ϕ in this expression by its vacuum expectation value (20.58), we obtain

$$\Delta\mathcal{L}_e = -\frac{1}{\sqrt{2}}\lambda_e v\, \bar{e}_L e_R + \text{h.c.} + \cdots. \tag{20.99}$$

This is a mass term for the electron. The size of the mass is set by the vacuum expectation value of ϕ, rescaled by the new dimensionless coupling:

$$m_e = \frac{1}{\sqrt{2}}\lambda_e v. \tag{20.100}$$

Since the electron mass is proportional to v, one might expect that the masses of the electron and the W boson should be of the same order. In fact, taking the observed values, $m_e/m_W \sim 6 \times 10^{-6}$. Since λ_e is a renormalizable coupling, it must be treated as an input to the theory. Thus the GWS theory allows the electron to be very light, but it cannot explain why the electron is so light compared to the W boson.

We can write mass terms for the quark fields in the same way. Notice that, in the following expression, both terms are invariant under $SU(2)$ and have zero net Y:

$$\Delta\mathcal{L}_q = -\lambda_d \overline{Q}_L \cdot \phi\, d_R - \lambda_u \epsilon^{ab}\overline{Q}_{La}\phi_b^\dagger u_R + \text{h.c.} \tag{20.101}$$

Substituting the vacuum expectation value of ϕ from Eq. (20.58), these terms become

$$\Delta\mathcal{L}_q = -\frac{1}{\sqrt{2}}\lambda_d v\, \bar{d}_L d_R - \frac{1}{\sqrt{2}}\lambda_u v\, \bar{u}_L u_R + \text{h.c.} + \cdots, \tag{20.102}$$

standard mass terms for the d and u quarks. The GWS theory thus gives the relations

$$m_d = \frac{1}{\sqrt{2}}\lambda_d v, \qquad m_u = \frac{1}{\sqrt{2}}\lambda_u v. \tag{20.103}$$

As with the electron, the theory parametrizes but does not explain the small values of the d and u quark masses observed experimentally.

When additional generations of quarks are introduced into the theory, there can be additional coupling terms that mix generations. Alternatively, we can diagonalize the Higgs couplings by choosing a new basis for the quark fields. We will show that this is always possible in Section 20.3. However, this simplification of the Higgs couplings causes a complication in the gauge couplings. Let

$$u_L^i = (u_L, c_L, t_L)\ , \qquad d_L^i = (d_L, s_L, b_L) \tag{20.104}$$

denote the up- and down-type quarks in their original basis, and let $u_L'^i$ and $d_L'^i$ denote the quarks in the basis that diagonalizes their Higgs couplings. This latter basis is the physical one, since it is the basis that diagonalizes the mass matrix. The two bases are related by unitary transformations:

$$u_L^i = U_u^{ij} u_L'^j, \qquad d_L^i = U_d^{ij} d_L'^j. \tag{20.105}$$

In this new basis, the W boson current takes the form

$$J_W^{\mu+} = \frac{1}{\sqrt{2}}\bar{u}_L^i \gamma^\mu d_L^i = \frac{1}{\sqrt{2}}\bar{u}_L'^i \gamma^\mu (U_u^\dagger U_d)_{ij} d_L'^j. \tag{20.106}$$

This expression is conventionally written

$$J_W^{\mu+} = \frac{1}{\sqrt{2}}\bar{u}_L'^i \gamma^\mu V_{ij} d_L'^j, \tag{20.107}$$

where V_{ij} is a unitary matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The off-diagonal terms in V_{ij} allow weak-interaction transitions between quark generations. For example, restricting to two generations for simplicity and writing

$$V_{1j} d_L'^j = \cos\theta_c d_L' + \sin\theta_c s_L', \tag{20.108}$$

the term proportional to $\sin\theta_c$ allows an s quark to decay weakly to a u quark. We have made use of this structure in our discussion of the effective Lagrangian for K meson decays in Section 18.2. We will discuss CKM flavor mixing and its symmetry properties in more detail in Section 20.3.

It is interesting to note that there is no term within the structure we have described that gives a mass to the neutrino. If we wanted to generalize Eq. (20.98) to allow a neutrino mass term, we would have to introduce a new

fermion field ν_R that is completely neutral under $SU(2)\times U(1)$. Then we could write the Higgs coupling

$$\Delta\mathcal{L}_\nu = -\lambda_\nu \epsilon^{ab}\overline{E}_{La}\phi_b^\dagger \nu_R + \text{h.c.}, \tag{20.109}$$

which would give the ν_e a mass, presumably comparable to that of the electron. But we know from experiment that neutrino masses are extremely small; the mass of the ν_e is known to be less than 10 eV. This extreme suppression of the neutrino masses would be naturally explained if the states ν_R do not exist. We will show in Section 20.3 that this assumption also implies that there are no transitions between leptons of different generations; this result is also in accord with very strong experimental bounds.

The Higgs Boson

This discussion of fermion mass generation emphasizes that the scalar field that causes spontaneous breaking of the gauge symmetry is an important ingredient in the structure of the Glashow-Weinberg-Salam theory. We should therefore ask whether it has any more direct manifestations.

To investigate this question, we will work in the *unitarity gauge*, analogous to that used for the Abelian model in Eq. (20.12). Let us parametrize the scalar field ϕ by writing

$$\phi(x) = U(x)\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{20.110}$$

The two-component spinor on the right has an arbitrary real-valued lower component, given by the vacuum expectation value of ϕ plus a fluctuating real-valued field $h(x)$ with $\langle h(x)\rangle = 0$. This spinor is acted on by a general $SU(2)$ gauge transformation $U(x)$ to produce the most general complex-valued two-component spinor. We can now make a gauge transformation to eliminate $U(x)$ from the Lagrangian. This reduces ϕ to a field with one physical degree of freedom.

An explicit renormalizable Lagrangian that leads to a vacuum expectation value for ϕ is

$$\mathcal{L} = \left|D_\mu \phi\right|^2 + \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2. \tag{20.111}$$

The minimum of the potential energy occurs at

$$v = \left(\frac{\mu^2}{\lambda}\right)^{1/2}. \tag{20.112}$$

In the unitarity gauge, the potential energy terms in (20.111) take the form

$$\begin{aligned}\mathcal{L}_V &= -\mu^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 \\ &= -\frac{1}{2}m_h^2 h^2 - \sqrt{\frac{\lambda}{2}}\, m_h h^3 - \frac{1}{4}\lambda h^4.\end{aligned} \tag{20.113}$$

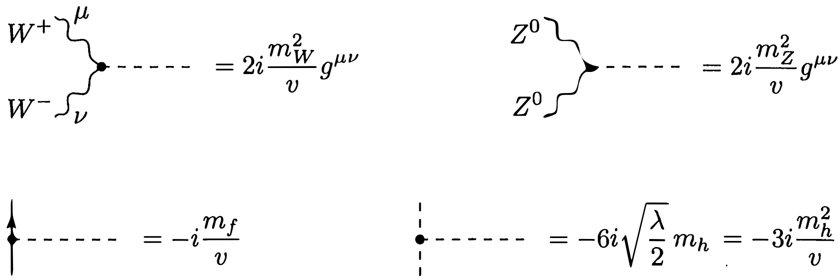


Figure 20.6. Feynman rules for the couplings of the Higgs boson to vector bosons, to fermions, and to itself.

The quantum of the field $h(x)$ is a scalar particle with mass

$$m_h = \sqrt{2}\mu = \sqrt{2\lambda}\,v. \tag{20.114}$$

This particle is known as the *Higgs boson.* As for the fermions in the GWS theory, the Higgs boson has a mass whose general magnitude is controlled by the vacuum expectation value v, but whose precise value is determined by a new, unspecified, renormalizable coupling constant.

The expansion of the kinetic energy term of (20.111) in unitarity gauge yields the gauge boson mass term (20.62), plus additional terms involving the Higgs boson field. Explicitly,

$$\mathcal{L}_K = \frac{1}{2}(\partial_\mu h)^2 + \left[m_W^2 W^{\mu+}W_\mu^- + \frac{1}{2}m_Z^2 Z^\mu Z_\mu\right]\cdot\left(1+\frac{h}{v}\right)^2, \tag{20.115}$$

where m_W and m_Z are given by Eqs. (20.63).

Finally, the fermion mass terms in Eqs. (20.98) and (20.101) lead to direct couplings of the Higgs boson to fermions. Evaluating these terms in unitarity gauge, we find that, for any quark or lepton flavor, the Higgs boson couples according to

$$\mathcal{L}_f = -m_f\bar{f}f\left(1+\frac{h}{v}\right). \tag{20.116}$$

The Higgs boson couplings in Eqs. (20.113), (20.115), and (20.116) lead to the Feynman rules shown in Fig. 20.6.

In general, the couplings of the Higgs boson to other particles of the weak interaction theory are proportional to the masses of those particles. Thus, the particles that are most easily made in the laboratory have very weak couplings to the Higgs boson, which makes it difficult to observe this particle. In any event, the Higgs boson has not yet been found. As of this writing, the Higgs boson that we have just described has been searched for and excluded for values of m_h below 60 GeV. If the self-coupling λ is large, however, the Higgs boson could have a mass as large as 1000 GeV; thus, a large dynamic range remains unexplored.

The phenomenological properties of the Higgs boson are worked out in more detail in the Final Project of Part III.

A Higgs Sector?

Since there is no experimental evidence for the existence of the simple Higgs boson contained in the GWS model, it is worth asking whether the W and Z bosons might acquire mass by a more complicated mechanism. There are two aspects to this question.

First, is it certain that the W and Z bosons are gauge bosons of a spontaneously broken $SU(2) \times U(1)$ symmetry? The evidence for this idea comes from the universality of the couplings of the various quarks and leptons to the weak interactions. This universality is tested in the fact that the same value of the Fermi constant describes all charged-current weak-interaction processes, and that this same strength of coupling combined with a single value of $\sin^2 \theta_w$ describes the whole range of weak neutral current phenomena. We have seen, especially in the discussion of Chapter 16, that the principle of local gauge invariance leads naturally to the prediction of universal, flavor-independent, coupling constants. No other principle is known that would explain this striking regularity. Thus there is compelling evidence that the underlying theory of the weak interactions is a spontaneously broken gauge theory.

However, it is quite possible that the mechanism of the spontaneous breaking of $SU(2) \times U(1)$ is more complicated than the simple model of a single scalar field that we have written in Eq. (20.111). In principle, the breaking of $SU(2) \times U(1)$ might be the result of the dynamics of a complicated new set of particles and interactions, which we will refer to as the *Higgs sector*. Experiment gives us only three properties of this new sector: First, it must generate the masses of the quarks and leptons. Second, it must generate the masses of the W and Z bosons. The third piece of information, which is the only nontrivial one, comes from the relation (20.73) between weak boson masses in the GWS theory,

$$m_W = m_Z \cos \theta_w. \tag{20.117}$$

This relation is satisfied experimentally to better than 1% accuracy, that is, to the level of one-loop radiative corrections. Whatever complicated mechanism we invoke to generate the spontaneous breaking of $SU(2) \times U(1)$, it should reproduce this relation in a natural way.

To understand the implications of relation (20.117), we must analyze the gauge boson mass matrix without assuming that $SU(2) \times U(1)$ is broken by the expectation value of a scalar field. Actually, it is possible to compute the gauge boson mass matrix under much less restrictive assumptions, using the argument given at the very end of Section 20.1. There we constructed the gauge boson mass matrix from the matrix elements for gauge currents to create or destroy Goldstone bosons. We will now show that relation (20.117) follows for a large class of models of $SU(2) \times U(1)$ breaking for which these matrix elements satisfy certain simple properties.

Any model of weak-interaction symmetry breaking must contain some set of fields that is responsible for the spontaneous breaking of $SU(2) \times U(1)$. Think of this sector of the theory as a field theory with a global $SU(2) \times U(1)$ symmetry, which is promoted to a local symmetry through its coupling to gauge bosons. In the theory with global symmetry, this symmetry is spontaneously broken to $U(1)$. Since three continuous symmetries are spontaneously broken, this sector must supply three Goldstone bosons, which will eventually be eaten by W^+, W^-, and Z^0. Call these three bosons π_a, where $a = 1, 2, 3$. Let $J^{\mu a}$ be the $SU(2)$ symmetry currents of the new sector, and let $J^{\mu Y}$ be the $U(1)$ current. The gauge boson mass matrix will then be constructed from the matrix elements (20.46), which here take the form

$$\langle 0| J^{\mu A} |\pi_b(p)\rangle = -ip^\mu F^A{}_b, \tag{20.118}$$

with $A = 1, 2, 3, Y$ and $b = 1, 2, 3$. Using the method of Eq. (20.55), we find that the gauge boson vacuum polarization contains the pole term

$$-\frac{i}{k^2}(g_A F^A{}_c)(g_B F^B{}_c), \tag{20.119}$$

summed over c, where $g_A = g$ for $A = 1, 2, 3$ and $g_A = g'$ for $A = Y$. Then we can identify the gauge boson mass matrix as

$$m^2_{AB} = g_A g_B F^A{}_c F^B{}_c. \tag{20.120}$$

To reproduce the known form of the weak gauge boson mass matrix, we must now place constraints on the $F^A{}_b$. First we must insure that the photon remains massless. This follows if the linear combination of charges (20.69) annihilates the vacuum. In the language of Eq. (20.118), we must insist that the corresponding linear combination of currents cannot excite a Goldstone boson:

$$\langle 0| \left(J^{\mu 3} + J^{\mu Y}\right) |\pi^a(p)\rangle = 0. \tag{20.121}$$

We can also achieve relation (20.117), using the following additional assumption: The symmetry-breaking sector has an $SU(2)$ global symmetry, under which the three Goldstone bosons and the three $SU(2)$ gauge currents transform as triplets, which remains exact when the $SU(2)$ gauge symmetry is spontaneously broken. This global $SU(2)$ symmetry implies that, if $A = a = 1, 2, 3$ in Eq. (20.118),

$$\langle 0| J^{\mu a} \left|\pi^b(p)\right\rangle = -iFp^\mu \delta^{ab}, \tag{20.122}$$

where F is a parameter with the dimensions of mass. Combining (20.122) and (20.121), we have

$$\langle 0| J^{\mu Y} \left|\pi^3(p)\right\rangle = +iFp^\mu. \tag{20.123}$$

Inserting this form for $F^A{}_b$ into (20.120), we find the gauge boson mass matrix

$$m^2 = F^2 \begin{pmatrix} g^2 & & & \\ & g^2 & & \\ & & g^2 & -gg' \\ & & -gg' & g'^2 \end{pmatrix}, \tag{20.124}$$

where the matrix acts on the gauge boson $(A_\mu^1, A_\mu^2, A_\mu^3, B_\mu)$. The eigenvectors of this matrix are precisely (20.63) and (20.64). To reproduce the eigenvalues, we need only define

$$v = 2F. \tag{20.125}$$

We have now shown that the GWS relation between the W and Z boson masses is not special to the situation in which the gauge symmetry is broken by a single scalar field. This relation follows from the much more general assumption of an unbroken global $SU(2)$ symmetry of the Higgs sector. This symmetry is often called *custodial $SU(2)$ symmetry.** We have seen this symmetry already as the global $SU(2)$ symmetry of the weak-interaction effective Lagrangian (20.94).

For the case of a single scalar field, the custodial symmetry arises in the following way: If we write the field ϕ in terms of its four real components, the Lagrangian (20.111) (ignoring the gauge couplings) has $O(4)$ global symmetry. The vacuum expectation value of ϕ breaks this symmetry down to $O(3)$, that is, $SU(2)$.

However, there are many other quantum field theories that break $SU(2)$ spontaneously while leaving another global $SU(2)$ symmetry unbroken. One rather complex example is given by QCD with two massless flavors, if we identify the gauged $SU(2)$ with the symmetry generated by U_L in (19.82) and identify the custodial $SU(2)$ with vectorial isospin symmetry. A copy of the familiar strong interactions with a mass scale large enough to give $F = 125$ GeV would be a perfectly acceptable model for the Higgs sector. (Unfortunately, it is not easy in this model to generate masses for the quarks and leptons.)

The question of the nature of the Higgs sector and the explicit mechanism of $SU(2) \times U(1)$ breaking is probably the most pressing open problem in the theory of elementary particles. We will discuss this question further in the Epilogue.

20.3 Symmetries of the Theory of Quarks and Leptons

Putting together the theory of strong interactions described in Chapter 17 and the theory of weak and electromagnetic interactions described in the previous section, we have now constructed a complete description of elementary particle interactions. It is interesting to investigate the symmetries of this theory, to ask what might be the fundamental symmetries of the underlying description of Nature.

We have already seen, in the arguments leading up to Eq. (15.17), that the Lagrangian of a gauge theory is highly restricted by the conditions of renormalizability and gauge invariance. In this section, we will construct the

*P. Sikivie, L. Susskind, M. Voloshin, and V. Zakharov, *Nucl. Phys.* **B173**, 189 (1980).

most general renormalizable Lagrangian consistent with the $SU(3) \times SU(2) \times U(1)$ gauge symmetries of the strong, weak and electromagnetic interactions. We can then ask what further global symmetries we must impose on this theory in order to give it the global symmetries that we see in Nature.

As a first step, we will ignore the Higgs scalar field and the mass terms of quarks, leptons, and gauge bosons. Then the Lagrangian of the theory of quarks and leptons is entirely specified by gauge invariance and renormalizability. We have

$$\mathcal{L}_K = -\frac{1}{4}\sum_i (F^a_{i\mu\nu})^2 + \sum_J \bar{\psi}_J (i\not{D})\psi_J, \tag{20.126}$$

where the index i runs over the three factors of the gauge group and the index J runs over the various multiplets of chiral fermions.

In principle, we could add to (20.126) the following pseudoscalar pure gauge operators:

$$\Delta\mathcal{L}_\theta = \sum_i \frac{\theta_i g_i^2}{64\pi^2}\epsilon^{\mu\nu\lambda\sigma} F^a_{i\mu\nu} F^a_{i\lambda\sigma}. \tag{20.127}$$

These terms are apparently odd under both P and T. However, we saw at the end of Section 19.2 that terms of this form can be generated or canceled by making a change of variables in the effective action. For example, the change of variables on the right-handed electron field

$$e_R \rightarrow e^{i\alpha} e_R \tag{20.128}$$

produces, according to (19.78) or (19.79), a correction to the Lagrangian involving the P- and T-odd combination of field strengths for the $U(1)$ gauge field

$$\Delta\mathcal{L} = \alpha \cdot \frac{g'^2}{32\pi^2}\epsilon^{\mu\nu\lambda\sigma} F_{\mu\nu} F_{\lambda\sigma}. \tag{20.129}$$

The coefficient of (20.129) differs from the corresponding coefficient in (19.79) because we transform only the right-handed chiral component of the electron field. If we were to transform another fermion field, of hypercharge Y, we would find a similar shift, with the coefficient proportional to Y^2. If this new field coupled to the $SU(2)$ or $SU(3)$ gauge fields, we would also find terms proportional to those field strengths. Thus, we can eliminate the term in (20.127) involving the $U(1)$ field strengths by making the change of variables (20.128) with $\alpha = -\frac{1}{2}\theta_1$. We can eliminate all three terms in (20.127) by making appropriate chiral rotations on three fermion multiplets, say, e_R, E_L, and Q_L. The change of variables (20.128), which rotates the right-handed electron field, is not symmetric under parity and, in fact, changes the definition of the parity operation. By making this change of variables, we are choosing new coordinates in which the P and T transformation properties of the whole theory are as simple as possible.

Let us now investigate the properties of the Lagrangian (20.126) under P, C, and T. The couplings of the QCD gauge bosons are invariant to each

of these symmetries separately. However, the couplings of the $SU(2)$ gauge bosons violate P and C as much as possible. Recall from Section 3.6 that P converts a left-handed electron to a right-handed electron, and that C converts a left-handed electron to a left-handed positron. Each of these operations converts a particle that couples to $SU(2)$ gauge bosons to one that does not. However, the combination of these two operations interchanges left-handed particles with right-handed antiparticles. Thus the combined operation CP is a symmetry of (20.126). This Lagrangian is also invariant under time reversal.

Thus, we see that the discrete symmetries of C and P, on the one hand, and CP and T, on the other, stand on a very different footing in gauge field theories. Any chiral gauge theory will naturally violate C and P. At this level in our analysis, it is a mystery why C and P should be observed to be approximate symmetries of Nature. On the other hand, every theory of gauge bosons and massless fermions respects CP and T. It is known experimentally that Nature contains some interaction that violates CP, since the CP selection rules are weakly violated in the decays of the K^0 meson. But to find a source for this violation, we must add terms to our basic gauge theory (20.126).

We must, first of all, add dynamics to (20.126) that will cause the spontaneous breaking of $SU(2) \times U(1)$. We will begin by working with the simplest model with one Higgs scalar field ϕ. The most general renormalizable Lagrangian for ϕ is

$$\mathcal{L}_\phi = \left|D_\mu \phi\right|^2 + \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \tag{20.130}$$

The Hermiticity of $\mathcal{L}_\phi$ implies that the parameters μ^2 and λ are real. Thus this Lagrangian respects P, C, and T. As discussed at the end of the previous section, this Lagrangian also automatically has the custodial $SU(2)$ symmetry required to produce the mass relation (20.117).

Finally, we must add the terms that couple the Higgs field to the quarks and leptons. Here, renormalizability and gauge invariance provide the weakest constraints, and there are many allowed interactions. We will first analyze the coupling of ϕ to the quark fields, and then generalize this discussion to the leptons.

In writing the Higgs field couplings to the quarks, we should recall that there are known to be three generations of quarks and leptons. Thus there are three doublets of left-handed quarks:

$$Q_L^i = \begin{pmatrix} u^i \\ d^i \end{pmatrix}_L = \left(\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \right). \tag{20.131}$$

There are six right-handed quarks, three with $Y = \frac{2}{3}$ and three with $Y = -\frac{1}{3}$:

$$u_R^i = \left(u_R, c_R, t_R\right), \qquad d_R^i = \left(d_R, s_R, b_R\right). \tag{20.132}$$

When we couple gauge fields to these quarks, we replace the ordinary derivatives with covariant derivatives. This automatically gives all of the quarks the same coupling to QCD and all quarks of the same type the same coupling to

the weak interactions. It does not allow mixing between the various quark flavors. However, the coupling of the Higgs field to the quarks does not follow from a gauge principle and so need not have any of these restrictions. Unless we require quark flavor conservation by postulating a new discrete symmetry of the theory, the Higgs couplings will, in general, mix the various flavors.

If we do not impose any additional symmetries on the theory, we must write the most general renormalizable gauge-invariant coupling with the structure of Eq. (20.101):

$$\mathcal{L}_m = -\lambda_d^{ij}\overline{Q}_L^i \cdot \phi\, d_R^j - \lambda_u^{ij}\epsilon^{ab}\overline{Q}_{La}^i\phi_b^\dagger u_R^j + \text{h.c.}, \tag{20.133}$$

where λ_d^{ij} and λ_u^{ij} are general, not necessarily symmetric or Hermitian, complex-valued matrices. The operation of CP interchanges the operators written in (20.133) with their Hermitian conjugates without changing the coefficients; thus, CP is equivalent to the substitutions

$$\lambda_d^{ij} \to (\lambda_d^{ij})^*, \qquad \lambda_u^{ij} \to (\lambda_u^{ij})^*. \tag{20.134}$$

CP would be a symmetry of (20.133) if the matrices λ^{ij} were real-valued; however, there is no principle that requires this. Without the imposition of further symmetry requirements, it seems that (20.133) does maximum violence to all discrete and flavor conservation symmetries.

However, just as we were able to eliminate the T-violating terms (20.127) by making a chiral rotation, we can simplify the form of (20.133) by appropriate chiral transformations. To find the required transformations, diagonalize the Hermitian matrices obtained by squaring λ_d and λ_u. Define unitary matrices U_u and W_u by

$$\lambda_u\lambda_u^\dagger = U_u D_u^2 U_u^\dagger, \qquad \lambda_u^\dagger\lambda_u = W_u D_u^2 W_u^\dagger, \tag{20.135}$$

where D_u^2 is a diagonal matrix with positive eigenvalues. Then

$$\lambda_u = U_u D_u W_u^\dagger, \tag{20.136}$$

where D_u is the diagonal matrix whose diagonal elements are the positive square roots of the eigenvalues of (20.135). We can define unitary matrices U_d and W_d in a similar way and decompose λ_d as

$$\lambda_d = U_d D_d W_d^\dagger. \tag{20.137}$$

Now make the change of variables

$$u_R^i \to W_u^{ij} u_R^j, \qquad d_R^i \to W_d^{ij} d_R^j. \tag{20.138}$$

This eliminates the unitary matrices W_u and W_d from the Higgs coupling (20.133). Since each of the three u_R^i and each of the three d_R^i have the same coupling to the gauge fields, W_u and W_d commute with the corresponding covariant derivatives. Thus, under (20.138),

$$\sum_i \left(\bar{u}_R^i(i\not{D})u_R^i + \bar{d}_R^i(i\not{D})d_R^i\right) \to \sum_i \left(\bar{u}_R^i(i\not{D})u_R^i + \bar{d}_R^i(i\not{D})d_R^i\right), \tag{20.139}$$

and so W_u and W_d disappear from the theory.

The analogous transformation on the left-handed fields also makes a dramatic simplification. Make the change of variables

$$u_L^i \to U_u^{ij} u_L^j, \qquad d_L^i \to U_d^{ij} d_L^j. \tag{20.140}$$

This transformation eliminates U_u, U_d from the terms in (20.133) that involve the lower component of the Higgs field. In unitarity gauge, only these terms survive. By combining the diagonal elements of D_u and D_d with the vacuum expectation value of the Higgs field, we can relate these elements to quark masses:

$$m_u^i = \frac{1}{\sqrt{2}} D_u^{ii} v, \qquad m_d^i = \frac{1}{\sqrt{2}} D_d^{ii} v. \tag{20.141}$$

With this replacement, (20.133) takes the form

$$\mathcal{L}_m = -m_d^i \bar{d}_L^i d_R^i \Big(1 + \frac{h}{v}\Big) - m_u^i \bar{u}_L^i u_R^i \Big(1 + \frac{h}{v}\Big) + \text{h.c.} \tag{20.142}$$

This has the standard form of quark mass terms and Higgs boson couplings. The transformations (20.138) and (20.140) thus convert the quark fields to the basis of mass eigenstates. In this basis, the mass terms and Higgs couplings are diagonal in flavor and conserve P, C, and T.

Since left-handed u and d quarks have identical couplings to QCD, the matrices U_u and U_d commute with the QCD couplings in the covariant derivative. However, u_L and d_L are mixed by the weak interactions, and so we must investigate the effect of (20.140) on the $SU(2)\times U(1)$ couplings more carefully. This is most easily done by referring to the Lagrangian (20.79). The matrices U_u and U_d cancel out of the pure kinetic terms in the first line of (20.79). They also cancel out of the electromagnetic current J^μ_{EM}; for example,

$$\bar{u}_L^i \gamma^\mu u_L^i \to \bar{u}_L^i U_u^{\dagger ij} \gamma^\mu U_u{}^{jk} u_L^k = \bar{u}_L^i \gamma^\mu u_L^i. \tag{20.143}$$

By the same logic, U_u and U_d cancel out of the Z^0 boson current.

However, in the current that couples to the W boson field, we find

$$J^{\mu+} = \frac{1}{\sqrt{2}} \bar{u}_L^i \gamma^\mu d_L^i \to \frac{1}{\sqrt{2}} \bar{u}_L^i \gamma^\mu \left(U_u^\dagger U_d\right)^{ij} d_L^j. \tag{20.144}$$

That is, the charge-changing weak interactions link the three u_L^i quarks with a unitary rotation of the triplet of d_L^i quarks, with this rotation given by the unitary matrix

$$V = U_u^\dagger U_d. \tag{20.145}$$

The matrix V is known as the *Cabibbo-Kobayashi-Maskawa* (CKM) mixing matrix.

The matrix V can have complex elements, but we can remove phases from V by performing phase rotations of the various quark fields. Before analyzing the case of three generations, it is useful to consider the case of two

generations—u, d, c, and s. In this case, V is a 2×2 unitary matrix. Such a matrix has 4 parameters; we can write its most general form as

$$V = \begin{pmatrix} \cos\theta_c e^{i\alpha} & \sin\theta_c e^{i\beta} \\ -\sin\theta_c e^{i(\alpha+\gamma)} & \cos\theta_c e^{i(\beta+\gamma)} \end{pmatrix}. \tag{20.146}$$

One parameter of V is a rotation angle, and the other three are phases. We can remove these phases by performing the change of variables on the quark fields

$$q_L^i \to \exp[i\alpha^i] q_L^i. \tag{20.147}$$

This global phase rotation has no effect on any term of the Lagrangian except for the weak charged current (20.144). A phase rotation that is equal for all four quark flavors cancels out of (20.144). However, the other three possible phase transformations are just what we need to eliminate α, β, and γ.

When we have chosen the phases of the quark fields in this way, V takes the form

$$V = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix}. \tag{20.148}$$

Then the quark terms in the weak charged current can be written

$$J^{\mu+} = \frac{1}{\sqrt{2}}\big(\cos\theta_c \bar{u}_L\gamma^\mu d_L + \sin\theta_c \bar{u}_L\gamma^\mu s_L - \sin\theta_c \bar{c}_L\gamma^\mu d_L + \cos\theta_c \bar{c}_L\gamma^\mu s_L\big). \tag{20.149}$$

We have already seen, in Eqs. (18.31) and (18.32), that this is the way the s quark enters the weak interactions. The angle θ_c is the Cabibbo angle, as defined in Eq. (18.30).

The same set of arguments can be made for the theory with three generations. Here V is a general unitary 3×3 matrix. Such a matrix has 9 parameters. Of these, 3 are rotation angles; this is the number of parameters of an $O(3)$ rotation. The remaining 6 parameters are phases. We can remove these phases by making phase rotations of quark fields as in (20.147), but the overall phase is redundant, so we can remove only 5 of these phases. The final form of V contains 3 angles, of which one is the Cabibbo angle, and one phase. After all the transformations we have made, this one phase that makes some couplings of the W^+ to quarks complex is the only remaining parameter that violates CP.

We began this argument from a Lagrangian for the quark-Higgs boson coupling that seemed to violate all possible flavor symmetries and all discrete spacetime symmetries. However, by making changes of variables on the fermion fields, we have been able to dramatically simplify the form of the Lagrangian. If we keep only those terms involving the massless gauge bosons, the photon and the gluons, plus the mass terms and interactions written in (20.142), we see that this set of terms conserves P, C, T, and all flavor symmetries. This dramatic simplification occurs because the unbroken gauge symmetry of Nature, the gauge symmetry of QCD and QED, is nonchiral and can be written as acting on Dirac fermions. Since we have omitted only the

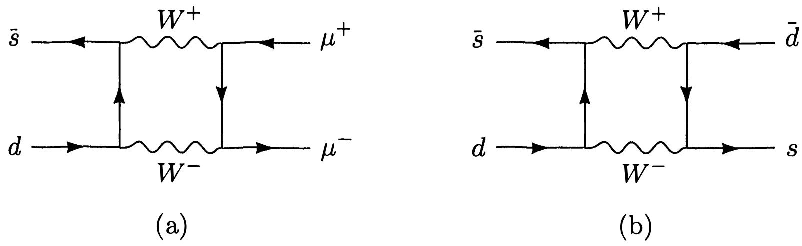


Figure 20.7. Higher-order diagrams that seem to give the leading contributions to flavor-changing weak neutral current processes: (a) $K^0 \to \mu^+\mu^-$; (b) $K^0 \leftrightarrow \overline{K}^0$.

effects mediated by the massive W and Z bosons, this much of the analysis already guarantees that Nature will appear, to a high degree of approximation, to respect the three separate discrete symmetries and all quark flavor conservation laws. Notice that we did not assume any fundamental global symmetries, but depended only on the assignment of gauge quantum numbers in the $SU(3) \times SU(2) \times U(1)$ gauge theory.

If we include the Z boson and the weak neutral current, we have a theory that violates P and C through Z exchange but that respects CP. In addition, this theory respects all flavor conservation laws. We describe this situation by saying that there is no flavor-changing weak neutral current. The experimental evidence for this statement is quite impressive. The best tests come from the study of the neutral K^0 meson, which is an $s\bar{d}$ bound state and so could decay by Z^0 exchange if this boson coupled to a flavor-changing current. In fact, the decay $K^0 \to \mu^+\mu^-$ is highly suppressed, to the level of the one-loop weak interaction correction shown in Fig. 20.7(a). Similarly, the interconversion of K^0 and $\overline{K}^0$, which could proceed directly if the Z^0 could change flavor, is suppressed to the level of the contribution shown in Fig. 20.7(b).

On the other hand, W bosons couple to currents that can change quark flavor, in a pattern parametrized by the Cabibbo angle and the other angles in the CKM matrix. Thus, heavy quark flavors decay by W boson exchange processes. Since the W couples to a current that contains only left-handed quarks, it mediates an interaction that violates P and C maximally. This violation of discrete symmetries is concealed from our ordinary experience because the amplitude for W exchange is small. However, this P and C violation is a dramatic qualitative feature of weak decays.

Since the coupling of the W to quarks contains an irreducible phase, these couplings in principle can violate CP. However, we have seen that this phase can be removed in a theory with only two generations. This means that the phase of the CKM matrix can have physical consequences only in a process that involves all three generations. Typically, this means that the CKM phase can contribute only to weak interaction loop corrections or to complicated exclusive decay processes. Thus the $SU(3) \times SU(2) \times U(1)$ theory can account

for CP violation, and also explains why this effect is much weaker even than the weak interactions. It is interesting to note that Kobayashi and Maskawa originally proposed the existence of the third generation in order to provide a mechanism for CP violation.†

On the other hand, at this moment there is no conclusive evidence that the origin of CP violation is indeed the phase of the CKM matrix. All of the arguments we have given in this section have used the simplest model of the Higgs sector, in which this sector consists of a single scalar field. More general models of the Higgs sector may leave behind a more complicated set of quark-Higgs couplings than appear in (20.142), and some of these may violate CP. In addition, there may be terms in the Higgs sector itself that lead to CP violation. The origin of the observed CP violation is still an open problem that needs both theoretical and experimental exploration.

Before leaving this subject, we must discuss one more aspect of this argument that is still mysterious. To simplify the Lagrangian of the gauge theory of quarks to its final form, we needed to make chiral changes of variables in the functional integral. We saw in Section 19.2, and we reviewed at the beginning of this section, that such changes of variables produce the new P- and T-violating terms written in Eq. (20.127). It can be shown, using the fact that these terms are total derivatives, that the terms involving $SU(2)$ and $U(1)$ field strengths have no observable effects. However, the term involving QCD field strengths can induce an electric dipole moment for the neutron, a T-violating effect that has been searched for and excluded at an impressive level of accuracy. Thus the P- and T-violating combination of QCD field strengths cannnot be allowed to appear in the Lagrangian. On the other hand, if the original up and down quark Higgs coupling matrices were of the most general possible form, it seems that this cannot be avoided. This problem is known as the *strong CP problem*. To solve this problem, one must either constrain the Higgs coupling matrices, violating the spirit of the argument we have just concluded, or one must add additional structure to the Higgs sector.‡

Finally, let us discuss the general form and simplification of the Higgs boson couplings to leptons. When we wrote the Glashow-Weinberg-Salam Lagrangian in the previous section, we noted that no gauge field coupled to the right-handed neutrino. Thus, we chose to eliminate this particle from the theory. We might need right-handed components of the neutrinos to construct neutrino mass terms, but at the moment there is no evidence for nonzero neutrino masses. Thus, in the remainder of this section, we will assume that there are no right-handed neutrinos and work out the consequences of this assumption.*

---

†M. Kobayashi and T. Maskawa, *Prog. Theor. Phys.* **49**, 652 (1973).

‡The strong CP problem, its proposed solutions, and their unexpected implications are reviewed by R. D. Peccei in *CP Violation,* C. Jarlskog, ed. (World Scientific, 1989).

*In generalizations of the $SU(2) \times U(1)$ model, neutrinos can acquire Majorana

Generalizing Eq. (20.133), we can write the most general coupling of a Higgs boson to three generations of leptons. Since there are no right-handed neutrinos, the only possible coupling is

$$\mathcal{L}_m = -\lambda_\ell^{ij}\overline{E}_L^i \cdot \phi\, e_R^j + \text{h.c.} \tag{20.150}$$

To diagonalize this coupling, represent λ_ℓ in the form

$$\lambda_\ell = U_\ell D_\ell W_\ell^\dagger, \tag{20.151}$$

and eliminate the matrices U_ℓ and W_ℓ by the changes of variables

$$e_L^i \to U_\ell^{ij} e_L^j, \qquad \nu_L^i \to U_\ell^{ij}\nu_L^j, \qquad e_R^i \to W_\ell^{ij} e_R^j. \tag{20.152}$$

Since we are now making the same change of variables on the two components of the weak doublet E_L^i, this change of variables commutes with the $SU(2)$ interactions in the covariant derivative. Thus the unitary matrices U_ℓ and W_ℓ completely disappear from the theory. The result is a theory of leptons that conserves CP exactly and also conserves the lepton number of each generation. This last result is very accurately tested experimentally. For example, there is no evidence for the generation-changing muon decay processes $\mu^- \to e^-\gamma$ or $\mu^- \to e^-e^-e^+$; the branching ratios for these processes are known to be below 10^{-10}.

We have seen, then, that the $SU(3) \times SU(2) \times U(1)$ gauge theory of quarks and leptons does an excellent job of accounting for the symmetries and conservation laws that are observed in elementary particle phenomena. It predicts which symmetries should be exact in Nature and which should be approximate. For approximate symmetries, it gives an accurate estimate of the level of symmetry violation. Most remarkably (except for the one issue of the strong CP problem), none of these predictions depend on any underlying global discrete or flavor symmetries in the fundamental equations. The global symmetries that we observe in Nature follow only from gauge invariance and the specific representation assignments that we made in constructing our gauge theory description.

---

mass terms that are naturally very small. These models also respect the constraints on lepton flavor mixing described in the next paragraph. For an introduction to these ideas on neutrino mass, see P. Ramond, in *Perspectives in the Standard Model,* R. K. Ellis, C. T. Hill, and J. D. Lykken, eds. (World Scientific, 1992).

Problems

20.1 Spontaneous breaking of $SU(5)$. Consider a gauge theory with the gauge group $SU(5)$, coupled to a scalar field Φ in the adjoint representation. Assume that the potential for this scalar field forces it to acquire a nonzero vacuum expectation value. Two possible choices for this expectation value are

$$\langle\Phi\rangle = A\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & -4 \end{pmatrix} \quad \text{and} \quad \langle\Phi\rangle = B\begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix}.$$

For each case, work out the spectrum of gauge bosons and the unbroken symmetry group.

20.2 Decay modes of the W and Z bosons.

(a) Compute the partial decay widths of the W boson into pairs of quarks and leptons. Assume that the top quark mass m_t is larger than m_W, and ignore the other quark masses. The decay widths to quarks are enhanced by QCD corrections. Show that the correction is given, to order α_s, by Eq. (17.9). Using $\sin^2\theta_w = 0.23$, find a numerical value for the total width of the W^+.

(b) Compute the partial decay widths of the Z boson into pairs of quarks and leptons, treating the quarks in the same way as in part (a). Determine the total width of the Z boson and the fractions of the decays that give hadrons, charged leptons, and invisible modes $\nu\bar{\nu}$.

20.3 $e^+e^- \to$ hadrons with photon-Z^0 interference

(a) Consider a fermion species f with electric charge Q_f and weak isospin I^3_L for its left-handed component. Ignore the mass of the f. Compute the differential cross section for the process $e^+e^- \to f\bar{f}$ in the standard electroweak model. Include the effect of the Z^0 width using the Breit-Wigner formula, Eq. (7.60). Plot the behavior of the total cross section as a function of CM energy through the Z^0 resonance, for u, d, and μ.

(b) Compute the forward-backward asymmetry for $e^+e^- \to f\bar{f}$, defined as

$$A^f_{FB} = \frac{\left(\int_0^1 - \int_{-1}^0\right) d\cos\theta (d\sigma/d\cos\theta)}{\left(\int_0^1 + \int_{-1}^0\right) d\cos\theta (d\sigma/d\cos\theta)},$$

as a function of center of mass energy.

(c) Show that, just on the Z^0 resonance, the forward-backward asymmetry is given by

$$A^f_{FB} = \frac{3}{4}A^e_{LR}A^f_{LR}.$$

(d) Show that the cross section at the peak of the Z^0 resonance is given by

$$\sigma_{\text{peak}} = \frac{12\pi}{m_Z^2}\frac{\Gamma(Z^0 \to e^+e^-)\Gamma(Z^0 \to f\bar{f})}{\Gamma_Z^2},$$

where Γ_Z is the total width of the Z^0. Notice that both the total width of the Z^0 and the peak height are affected by the presence of extra invisible decay modes. Compute the shifts in Γ_Z and σ_{peak} that would be produced by a hypothetical fourth neutrino species, and compare these shifts to the cross section measurements shown in Fig. 20.5.

20.4 Neutral-current deep inelastic scattering.

(a) In Eq. (17.35), we wrote formulae for neutrino and antineutrino deep inelastic scattering with $W^\pm$ exchange. Neutrinos and antineutrinos can also scatter by exchanging a Z^0. This process, which leads to a hadronic jet but no observable outgoing lepton, is called the *neutral current* reaction. Compute $d\sigma/dxdy$ for neutral current deep inelastic scattering of neutrinos and antineutrinos from protons, accounting for scattering from u and d quarks and antiquarks.

(b) Next, consider deep inelastic scattering from a nucleus A with equal numbers of protons and neutrons. For such a target, $f_u(x) = f_d(x)$, and similarly for antiquarks. Show that the formulae in part (a) simplify in such a situation. In particular, let R^ν, $R^{\bar\nu}$ be defined as

$$R^\nu = \frac{d\sigma/dxdy(\nu A \to \nu X)}{d\sigma/dxdy(\nu A \to \mu^- X)}, \qquad R^{\bar\nu} = \frac{d\sigma/dxdy(\bar\nu A \to \bar\nu X)}{d\sigma/dxdy(\bar\nu A \to \mu^+ X)}.$$

Show that R^ν and $R^{\bar\nu}$ are given by the following simple formulae:

$$\begin{aligned} R^\nu &= \frac{1}{2} - \sin^2\theta_w + \frac{5}{9}\sin^4\theta_w(1+r), \\ R^{\bar\nu} &= \frac{1}{2} - \sin^2\theta_w + \frac{5}{9}\sin^4\theta_w(1+\frac{1}{r}), \end{aligned}$$

where

$$r = \frac{d\sigma/dxdy(\bar\nu A \to \mu^+ X)}{d\sigma/dxdy(\nu A \to \mu^- X)}.$$

These formulae remain true when R^ν and $R^{\bar\nu}$ are redefined to be the ratios of neutral- to charged-current cross sections integrated over the region of x and y that is observed in a given experiment.

(c) By setting r equal to the observed value—say, $r = 0.4$—and varying $\sin^2\theta_w$, the relations of part (b) generate a curve in the plane of R^ν versus $R^{\bar\nu}$ that is known as *Weinberg's nose*. Sketch this curve. The observed values of R^ν, $R^{\bar\nu}$ lie close to this curve, near the point corresponding to $\sin^2\theta_w = 0.23$.

20.5 A model with two Higgs fields.

(a) Consider a model with two scalar fields ϕ_1 and ϕ_2, which transform as $SU(2)$ doublets with $Y = 1/2$. Assume that the two fields acquire parallel vacuum expectation values of the form (20.23) with vacuum expectation values v_1, v_2. Show that these vacuum expectation values produce the same gauge boson mass matrix that we found in Section 20.2, with the replacement

$$v^2 \to (v_1^2 + v_2^2).$$

(b) The most general potential function for a model with two Higgs doublets is quite complex. However, if we impose the discrete symmetry $\phi_1 \to -\phi_1$, $\phi_2 \to \phi_2$,

the most general potential is

$$V(\phi_1, \phi_2) = -\mu_1^2 \phi_1^\dagger \phi_1 - \mu_2^2 \phi_2^\dagger \phi_2 + \lambda_1 (\phi_1^\dagger \phi_1)^2 + \lambda_2 (\phi_2^\dagger \phi_2)^2 + \lambda_3 (\phi_1^\dagger \phi_1)(\phi_2^\dagger \phi_2) + \lambda_4 (\phi_1^\dagger \phi_2)(\phi_2^\dagger \phi_1) + \lambda_5 \big((\phi_1^\dagger \phi_2)^2 + \text{h.c.}\big).$$

Find conditions on the parameters μ_i and λ_i so that the configuration of vacuum expectation values required in part (a) is a locally stable minimum of this potential.

(c) In the unitarity gauge, one linear combination of the upper components of ϕ_1 and ϕ_2 is eliminated, while the other remains as a physical field. Show that the physical charged Higgs field has the form

$$\phi^+ = \sin\beta\, \phi_1^+ - \cos\beta\, \phi_2^+,$$

where β is defined by the relation

$$\tan\beta = \frac{v_2}{v_1}.$$

(d) Assume that the two Higgs fields couple to quarks by the set of fundamental couplings

$$\mathcal{L}_m = -\lambda_d^{ij} \overline{Q}_L^i \cdot \phi_1 d_R^j - \lambda_u^{ij} \epsilon^{ab} \overline{Q}_{La}^i \phi_{2b}^\dagger u_R^j + \text{h.c.}$$

Find the couplings of the physical charged Higgs boson of part (c) to the mass eigenstates of quarks. These couplings depend only on the values of the quark masses and $\tan\beta$ and on the elements of the CKM matrix.

Chapter 21

Quantization of Spontaneously Broken Gauge Theories

In Chapter 20 we saw that when a gauge symmetry is spontaneously broken, the gauge bosons acquire mass. This phenomenon allowed us to construct a realistic theory of the weak interactions. Up to this point, however, we have discussed spontaneously broken gauge theories only in a simplistic way. To isolate the physical degrees of freedom, we have used the device of going to the unitarity gauge. However, it is not at all clear what the rules of perturbation theory are in this gauge, or how the unitarity gauge constraint is maintained when we compute Feynman diagrams. We have also seen that the Goldstone bosons that are absorbed into the massive gauge bosons play an important role in formal arguments about these theories, so we would like to quantize these theories in a gauge that does not eliminate these particles from the beginning.

In this chapter we will address these problems, by carrying out the formal gauge-fixing of theories with spontaneously broken gauge symmetry using the Faddeev-Popov method. We will define a class of gauges, called the R_ξ gauges, almost all of which contain the Goldstone bosons of the original spontaneous symmetry breaking. These particles cancel the effects of other unphysical particles in the formalism to maintain the unitarity of the theory. These cancellations are a more intricate version of the cancellations between gauge and ghost degrees of freedom that we saw in Chapter 16. However, we will see in Section 21.2 that a theory does not forget that it contains Goldstone bosons and that, under some circumstances, the properties of the Goldstone bosons in the theory without gauge couplings can carry over to the theory with massive gauge bosons.

Finally, having defined the perturbation theory and clarified the role of the Goldstone bosons in spontaneously broken gauge theories, we will carry out some explicit loop calculations of interest in the theory of weak interactions. Here we will see applications of the ideas of Chapter 11, that a theory with spontaneously broken symmetry can be renormalized with the counterterms of the symmetric Lagrangian. In Section 21.3 we will show through some examples that this result applies with equal force to gauge theories, and that it endows the weak-interaction gauge theory with substantial predictive power.

21.1 The R_ξ Gauges

In our discussion of the low-energy effective Lagrangian for weak interactions, we proposed in Eq. (20.89) the following expression for the propagator of a massive gauge boson:

$$\langle A^\mu(p)A^\nu(-p)\rangle \stackrel{?}{=} \frac{-ig^{\mu\nu}}{p^2-m^2}. \tag{21.1}$$

This expression is a natural first guess, generalizing the Feynman-'t Hooft gauge. However, it is unsatisfactory in a number of ways.

The most important of these defects concerns the treatment of gauge boson polarization states. The propagator (21.1) contains four components, corresponding to the transverse, longitudinal, and timelike polarizations. We saw in Chapters 5 and 16 that, for massless gauge bosons, the unphysical longitudinal and timelike components cancel in computations. For a massive gauge boson, however, the longitudinal polarization state corresponds to a real physical particle; we do not want it to cancel. Expression (21.1) does not take this change into account.

An Abelian Example

To understand this and other formal problems that arise for gauge theories with spontaneously broken symmetry, we need to carefully redo the Faddeev-Popov quantization of these theories. To begin, we will quantize the spontaneously broken Abelian gauge theory introduced in Eq. (20.1):

$$\mathcal{L} = -\tfrac{1}{4}(F_{\mu\nu})^2 + \left|D_\mu\phi\right|^2 - V(\phi), \tag{21.2}$$

with $D_\mu = \partial_\mu + ieA_\mu$. Here $\phi(x)$ is a complex scalar field. However, it will be most convenient to analyze the model by writing ϕ in terms of its real components,

$$\phi = \frac{1}{\sqrt{2}}\left(\phi^1 + i\phi^2\right). \tag{21.3}$$

Then the infinitesimal local symmetry transformation is

$$\delta\phi^1 = -\alpha(x)\phi^2, \qquad \delta\phi^2 = \alpha(x)\phi^1, \qquad \delta A_\mu = -\frac{1}{e}\partial_\mu\alpha. \tag{21.4}$$

Let us assume that $V(\phi)$ forces the scalar field to acquire a vacuum expectation value: $\langle\phi^1\rangle = v$. Then we should change variables by a shift:

$$\phi^1(x) = v + h(x); \qquad \phi^2 = \varphi. \tag{21.5}$$

The field ϕ^2 or φ is the Goldstone boson. The Lagrangian (21.2) now takes the form

$$\mathcal{L} = -\tfrac{1}{4}(F_{\mu\nu})^2 + \tfrac{1}{2}\left(\partial_\mu h - eA_\mu\varphi\right)^2 + \tfrac{1}{2}\left(\partial_\mu\varphi + eA_\mu(v+h)\right)^2 - V(\phi). \tag{21.6}$$

This Lagrangian is still invariant under an exact local symmetry,

$$\delta h = -\alpha(x)\varphi, \qquad \delta\varphi = \alpha(x)(v+h), \qquad \delta A_\mu = -\frac{1}{e}\partial_\mu\alpha. \tag{21.7}$$

Thus, in order to define the functional integral over the variables (h, φ, A_μ), we must introduce Faddeev-Popov gauge fixing.

Starting from the functional integral

$$Z = \int \mathcal{D}A\mathcal{D}h\mathcal{D}\varphi\; e^{i\int \mathcal{L}[A,h,\varphi]}, \tag{21.8}$$

we can introduce a gauge-fixing constraint as we did in Section 9.4. Following the steps leading from Eq. (9.50) to Eq. (9.54), we find

$$Z = C\cdot \int \mathcal{D}A\mathcal{D}h\mathcal{D}\varphi\; e^{i\int \mathcal{L}[A,h,\varphi]}\,\delta\big(G(A,h,\varphi)\big)\det\Big(\frac{\delta G}{\delta\alpha}\Big), \tag{21.9}$$

where C is a constant proportional to the volume of the gauge group and $G(A,h,\varphi)$ is a gauge-fixing condition. Alternatively, we can introduce the gauge-fixing constraint as $\delta(G(x)-\omega(x))$ and integrate over $\omega(x)$ with a Gaussian weight, as in the derivation of Eq. (9.56). This gives

$$Z = C'\cdot \int \mathcal{D}A\mathcal{D}h\mathcal{D}\varphi\; \exp\Big[i\int d^4x\big(\mathcal{L}[A,h,\varphi] - \tfrac{1}{2}(G)^2\big)\Big]\det\Big(\frac{\delta G}{\delta\alpha}\Big). \tag{21.10}$$

The gauge-fixing function G is arbitrary, but we can simplify our formalism by choosing it appropriately.

An especially convenient choice of the gauge-fixing function is

$$G = \frac{1}{\sqrt{\xi}}\big(\partial_\mu A^\mu - \xi e v\varphi\big). \tag{21.11}$$

When we form G^2, the term quadratic in A_μ will provide the same gauge-dependent addition to the gauge field action that we saw in the derivation of Eqs. (9.58) and (16.29). In addition, the cross term between A_μ and φ is engineered to cancel the quadratic term of the form $\partial_\mu\varphi A^\mu$ coming from the third term of (21.6). With this choice, the quadratic terms of the gauge-fixed Lagrangian ($\mathcal{L} - \frac{1}{2}G^2$) are

$$\begin{aligned}\mathcal{L}_2 = &-\frac{1}{2}A_\mu\Big(-g^{\mu\nu}\partial^2 + \big(1-\frac{1}{\xi}\big)\partial^\mu\partial^\nu - (ev)^2 g^{\mu\nu}\Big)A_\nu \\ &+\frac{1}{2}(\partial_\mu h)^2 - \frac{1}{2}m_h^2 h^2 + \frac{1}{2}(\partial_\mu\varphi)^2 - \frac{\xi}{2}(ev)^2\varphi^2.\end{aligned} \tag{21.12}$$

The mass term for the h field comes from the expansion of $V(\phi)$, as in (20.6). The mass term for the gauge field comes from the Higgs mechanism, that is, from the third term of (21.6). Notice that the formalism also produces a mass for the Goldstone boson φ:

$$m_\varphi^2 = \xi(ev)^2 = \xi m_A^2. \tag{21.13}$$

The fact that this mass is gauge-dependent is a signal that the Goldstone boson is a fictitious field, which will not be produced in physical processes.

To complete the Faddeev-Popov quantization procedure, we must derive the Lagrangian of the ghosts. This Lagrangian depends on the gauge variation of G, which can be computed by inserting (21.7) into (21.11). We find

$$\frac{\delta G}{\delta\alpha} = \frac{1}{\sqrt{\xi}}\Big(-\frac{1}{e}\partial^2 - \xi ev(v+h)\Big). \tag{21.14}$$

The determinant of this operator can be accounted for by including a set of Faddeev-Popov ghosts with the Lagrangian,

$$\mathcal{L}_{\text{ghost}} = \bar{c}\Big[-\partial^2 - \xi m_A^2\big(1+\frac{h}{v}\big)\Big]c, \tag{21.15}$$

where $m_A = ev$ as in Eq. (21.13). Since this is an Abelian gauge theory, the ghost field does not couple directly to the gauge field. It does, however, couple to the physical Higgs field, so it cannot be completely ignored as in QED.

From the quadratic terms in the Lagrangians for A_μ, h, φ, and the ghosts, we can readily find the propagators for these fields. All four propagators are shown in Fig. 21.1. The only complicated case is that of the gauge field. The term in (21.12) involving A_μ involves an operator whose Fourier transform is

$$\begin{aligned} &g^{\mu\nu}k^2 - \big(1-\frac{1}{\xi}\big)k^\mu k^\nu - m_A^2 g^{\mu\nu} \\ &\qquad = \Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\Big)(k^2 - m_A^2) + \Big(\frac{k^\mu k^\nu}{k^2}\Big)\frac{1}{\xi}(k^2 - \xi m_A^2). \end{aligned} \tag{21.16}$$

The inverse of this matrix gives the A_μ field propagator:

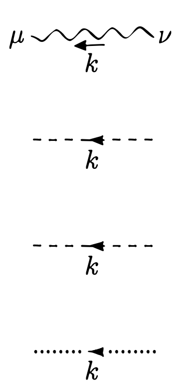
$$\begin{aligned} \langle A^\mu(k)A^\nu(-k)\rangle &= \frac{-i}{k^2-m_A^2}\Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\Big) + \frac{-i\xi}{k^2-\xi m_A^2}\Big(\frac{k^\mu k^\nu}{k^2}\Big) \\ &= \frac{-i}{k^2-m_A^2}\Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2-\xi m_A^2}(1-\xi)\Big). \end{aligned} \tag{21.17}$$

Notice that the transverse components of the A field and the component h of the Higgs field acquire the masses m_A, m_h that we found in Section 20.1. The unphysical components of A, the Goldstone bosons, and the ghosts all acquire the same gauge-dependent mass $\sqrt{\xi}m_A$.

ξ Dependence in Perturbation Theory

Because the parameter ξ was introduced only in the gauge fixing, we expect it to cancel out of all computations of expectation values of gauge-invariant operators and of S-matrix elements. This cancellation can be proved to all orders in perturbation theory by using the BRST symmetry of the gauge-fixed Lagrangian.* Here, however, we will simply illustrate the cancellation of ξ in a simple example.

*See, for example, Taylor (1976).



$$A_\mu: \quad = \frac{-i}{k^2 - m_A^2}\left(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2 - \xi m_A^2}(1-\xi)\right)$$

$$h: \quad = \frac{i}{k^2 - m_h^2}$$

$$\varphi: \quad = \frac{i}{k^2 - \xi m_A^2}$$

$$c: \quad = \frac{i}{k^2 - \xi m_A^2}$$

Figure 21.1. Propagators of the gauge field, Higgs fields, and ghosts in the Abelian model with spontaneously broken symmetry.

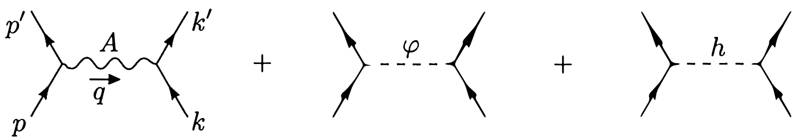


Figure 21.2. Diagrams contributing to fermion-fermion scattering at leading order in the Abelian model with spontaneous symmetry breaking.

Consider coupling a fermion to the spontaneously broken gauge theory through a chiral interaction:

$$\mathcal{L}_f = \bar{\psi}_L(i\not{D})\psi_L + \bar{\psi}_R(i\not{\partial})\psi_R - \lambda_f\left(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi^*\psi_L\right), \tag{21.18}$$

with $D_\mu = \partial_\mu + ieA_\mu$ as before. This is a stripped-down, Abelian version of the coupling of fermions to the weak interaction gauge theory. The fermion ψ receives a mass

$$m_f = \lambda_f \frac{v}{\sqrt{2}} \tag{21.19}$$

from the spontaneous symmetry breaking. (This theory has an axial vector anomaly that would render loop calculations inconsistent, but we will analyze it only at the level of tree diagrams.)

In this theory, the leading-order diagrams contributing to fermion-fermion scattering are those shown in Fig. 21.2. Notice that the contribution from the exchange of the unphysical particle φ must be included, since this particle appears in the Feynman rules. The ghosts do not appear in this process until the one-loop level. Since the propagator of the physical Higgs particle h is independent of ξ, the cancellation of the ξ dependence must take place between the transverse and longitudinal components of A_μ and the Goldstone boson φ.

The graph with exchange of the Goldstone boson has the value

$$i\mathcal{M}_\varphi = \left(\frac{\lambda_f}{\sqrt{2}}\right)^2 \bar{u}(p')\gamma^5 u(p) \frac{i}{q^2 - \xi m_A^2} \bar{u}(k')\gamma^5 u(k). \tag{21.20}$$

The ξ dependence of this expression must be canceled by that of the gauge boson exchange diagram,

$$\begin{aligned} i\mathcal{M}_A = (-ie)^2 \bar{u}(p')\gamma_\mu \Big(\frac{1-\gamma^5}{2}\Big) u(p) \\ \times \frac{-i}{q^2 - m_A^2}\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2 - \xi m_A^2}(1-\xi)\Big)\bar{u}(k')\gamma_\nu\Big(\frac{1-\gamma^5}{2}\Big)u(k). \end{aligned} \tag{21.21}$$

The ξ dependence of this term looks quite intricate. However, we can make some simplifications by rewriting the gauge boson propagator as

$$\begin{aligned} &\frac{-i}{q^2 - m_A^2}\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{m_A^2} + q^\mu q^\nu\Big[\frac{1}{m_A^2} - \frac{1}{q^2 - \xi m_A^2}(1-\xi)\Big]\Big) \\ &\qquad = \frac{-i}{q^2 - m_A^2}\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{m_A^2}\Big) + \frac{-i}{q^2 - \xi m_A^2}\Big(\frac{q^\mu q^\nu}{m_A^2}\Big). \end{aligned} \tag{21.22}$$

The first term of (21.22) is ξ-independent. The second term can be simplified in (21.21) by using the identity

$$\begin{aligned} q^\mu \bar{u}(p')\gamma_\mu\Big(\frac{1-\gamma^5}{2}\Big)u(p) &= \frac{1}{2}\bar{u}(p')\big[(\not{p} - \not{p}') - (\not{p} - \not{p}')\gamma^5\big]u(p) \\ &= \frac{1}{2}\bar{u}(p')\big[\not{p}'\gamma^5 + \gamma^5\not{p}\big]u(p) \\ &= m_f \bar{u}(p')\gamma^5 u(p), \end{aligned} \tag{21.23}$$

and the analogous identity on the other fermion line. After making these rearrangements and inserting the explicit values $m_f = \lambda_f v/\sqrt{2}$ and $m_A = ev$, the gauge boson exchange amplitude (21.21) takes the form

$$\begin{aligned} i\mathcal{M}_A =& (-ie)^2\bar{u}(p')\gamma_\mu\Big(\frac{1-\gamma^5}{2}\Big)u(p)\frac{i}{q^2 - m_A^2}\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{m_A^2}\Big)\bar{u}(k')\gamma_\nu\Big(\frac{1-\gamma^5}{2}\Big)u(k) \\ &+ \Big(\frac{\lambda_f}{\sqrt{2}}\Big)^2 \bar{u}(p')\gamma^5 u(p)\frac{-i}{q^2 - \xi m_A^2}\bar{u}(k')\gamma^5 u(k). \end{aligned} \tag{21.24}$$

The second term of (21.24) precisely cancels the Goldstone boson exchange diagram (21.20). The terms that remain in the fermion-fermion scattering amplitude are independent of ξ.

This demonstration merits two additional comments. First, throughout this book, we have become accustomed to dotting the gauge boson momentum into a gauge boson vertex and finding zero or contact terms. However, in spontaneously broken gauge theories, we typically find a different result: The fermionic current $\bar{\psi}\gamma^\mu(1-\gamma^5)\psi$ is not conserved, with the nonconservation being proportional to the fermion mass. This allows the manipulation (21.23) to contribute terms proportional to the Higgs boson vacuum expectation value,

which interplay with the Goldstone boson contributions. We will discuss this point further, and find a physical application of it, in Section 21.2.

The second point concerns the final form of the gauge-invariant sum of the gauge boson and Goldstone boson exchange diagrams. These give just the result we would have found by neglecting the Goldstone boson and computing the gauge boson exchange using the first term of (21.22) as the propagator:

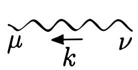
$$\langle A_\mu(q)A_\nu(-q)\rangle = \frac{-i}{q^2-m_A^2}\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{m_A^2}\Big). \tag{21.25}$$

The tensor structure represents a gauge boson polarization sum. To identify what vectors are summed over, notice that, if the vector boson is on-shell, and if we boost to its rest frame, this structure becomes precisely the projection onto the three purely spatial directions. These are the three polarization states of an on-shell massive vector particle. In a general frame, still for q^μ on-shell, the tensor in (21.25) remains the projection onto physical polarization states:

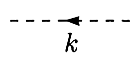
$$\sum_{\epsilon^\mu q_\mu=0} \epsilon^\mu\epsilon^{\nu *} = -\Big(g^{\mu\nu} - \frac{q^\mu q^\nu}{m_A^2}\Big). \tag{21.26}$$

Thus, in the cancellation of the ξ-dependent parts of the gauge boson propagator, we also find that the Goldstone boson diagram cancels the contribution of the unphysical timelike polarization state of the gauge boson, leaving over the required three physical polarizations.

The perturbation theory rules that we have developed have a very different character for different values of ξ. Thus, it is even more true in the case of spontaneously broken symmetry that we can find different special simplifications by choosing different values of this gauge parameter. For $\xi = 0$, Lorentz gauge, the Goldstone boson is massless and has exactly the couplings it has in the ungauged model of symmetry breaking, while the gauge boson propagator is purely transverse:



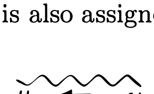
$$= \frac{-i}{k^2-m_A^2}\Big(g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\Big);$$



$$= \frac{i}{k^2}. \tag{21.27}$$

This gauge is especially useful for analyzing models of symmetry breaking. Both propagators have poles at $k^2 = 0$. However, we know that there are no corresponding physical particles, because these poles move away from $k^2 = 0$ as we change ξ, while the S-matrix must be ξ-independent.

For $\xi = 1$, we recover the simple form of the gauge boson propagator given in (21.1). This choice of the gauge boson propagator is not consistent, however, unless we also include Goldstone boson exchanges in which the Goldstone boson is also assigned the mass m_A:



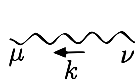
$$= \frac{-ig^{\mu\nu}}{k^2-m_A^2};$$

$$= \frac{i}{k^2-m_A^2}. \tag{21.28}$$

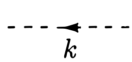
This gauge, still called the Feynman-'t Hooft gauge, is the most convenient one for general higher-order computations.

For any finite value of ξ, the gauge boson and Goldstone boson propagators fall off as $1/k^2$ and thus obey the general power-counting analysis of Section 10.1. It follows that, in any one of these gauges, the perturbation theory will be renormalizable, in the sense that the divergences are removed by a finite set of counterterms. Furthermore, the analysis of Section 11.6 tells us that the only counterterms required are those that are symmetric under the original global symmetry of the theory. However, we should require one further condition of our renormalization procedure: We should insist that the counterterms preserve local gauge invariance, and, in particular, preserve the property that S-matrix elements and the matrix elements of gauge-invariant operators are independent of ξ. This result was proved to all orders in perturbation theory by 't Hooft and Veltman and by Lee and Zinn-Justin.† Thus, in the gauge defined by any finite value of ξ, we can, in principle, straightforwardly compute a physical quantity to any order. The gauges defined by the possible values of ξ are known as the *renormalizability*, or R_ξ, gauges.

By taking the limit $\xi \to \infty$ of the R_ξ gauges, we find a gauge with very different simplifying features. In this limit, the unphysical degrees of freedom, which have masses proportional to $\sqrt{\xi}$, disappear from the theory. The gauge boson and Goldstone boson propagators become:



$$= \frac{-i}{k^2 - m_A^2}\left(g^{\mu\nu} - \frac{k^\mu k^\nu}{m_A^2}\right);$$



$$= 0. \qquad (21.29)$$

The gauge boson propagator contains exactly the three spacelike polarization states. In this gauge, the only singularities of Feynman diagrams correspond to the propagation of physical intermediate states. Thus, the unitarity of the S-matrix follows from the Cutkosky rules, as in the globally symmetric theories considered in Section 7.3, without the need to worry about the cancellation of unphysical states.‡ The $\xi \to \infty$ limit of the R_ξ gauges thus gives the quantum-mechanical realization of the *unitarity* (or U) gauge, introduced in Eq. (20.12).

It is not straightforward to prove renormalizability directly in the U gauge. In this gauge, the gauge boson propagator falls off more slowly than $1/k^2$ at large k. This signals trouble for the evaluation of loop diagrams. Typically, in fact, individual loop diagrams will diverge as $\log \xi$ or worse as $\xi \to \infty$. Still, the gauge invariance of the S-matrix implies that these divergences must cancel in the sum of all diagrams contributing to a given process, so that this sum has a smooth limit as $\xi \to \infty$. There is no difficulty of principle with the fact that we use one gauge to prove the renormalizability of spontaneously

†G. 't Hooft and M. J. G. Veltman, *Nucl. Phys.* **B50**, 318 (1972), B. W. Lee and J. Zinn-Justin, *Phys. Rev.* **D5**, 3121, 3137, 3155 (1972), **D7**, 1049 (1973).

‡In the more sophisticated language of Section 16.4, the crucial identity (16.54), which is required for the unitarity of the S-matrix, is true manifestly.

broken gauge theories and another gauge to prove their unitarity. In fact, this method of argumentation makes natural use of the underlying symmetries of the theory.

Non-Abelian Analysis

Now that we have thoroughly examined the R_ξ gauges for an Abelian gauge theory, we are ready to generalize to the non-Abelian case. There is no difficulty in being completely general, so let us consider a Yang-Mills gauge theory with gauge group G, spontaneously broken by the vacuum expectation value of a scalar field.

We will build on our classical analysis of this system following Eq. (20.13). As in that analysis, it will be most convenient to write the scalars as a multiplet ϕ_i of real-valued fields. Then the gauge transformation of the ϕ_i takes the form

$$\delta\phi_i = -\alpha^a(x)T^a_{ij}\phi_j, \tag{21.30}$$

where the T^a_{ij} are real, antisymmetric representation matrices of G. Similarly, the transformation of the gauge fields is

$$\delta A^a_\mu = \frac{1}{g}\partial_\mu\alpha^a - f^{abc}\alpha^b A^c_\mu = \frac{1}{g}(D_\mu\alpha)^a. \tag{21.31}$$

(If the gauge group is not simple, the coupling g need not be the same for every a.) The Lagrangian invariant under these gauge transformations is

$$\mathcal{L} = -\tfrac{1}{4}(F^a_{\mu\nu})^2 + \tfrac{1}{2}(D_\mu\phi)^2 - V(\phi), \tag{21.32}$$

with

$$D_\mu\phi_i = \partial_\mu\phi_i + gA^a_\mu T^a_{ij}\phi_j. \tag{21.33}$$

Assume that the potential $V(\phi)$ is minimized at a point where some of the components of ϕ acquire vacuum expectation values. As in (20.16), define

$$\langle\phi_i\rangle = (\phi_0)_i. \tag{21.34}$$

We will expand ϕ_i about this value:

$$\phi_i(x) = \phi_{0i} + \chi_i(x). \tag{21.35}$$

It will be convenient to divide the space of values χ_i into two subspaces. The vectors $T^a\phi_0$ correspond to symmetry transformations of the vacuum expectation value of ϕ. The field fluctuations along these directions are the Goldstone bosons. Let $\{n_i\}$ be an orthonormal basis for this subspace; then the unit vectors n_i are in 1-to-1 correspondence with the Goldstone bosons. The field fluctuations orthogonal to all of the vectors $T^a\phi_0$ correspond to the (massive) physical scalar fields of the spontaneously broken gauge theory.

In the discussion to follow, the vectors $T^a\phi_0$ will play an important role. We should then recall the notation for these vectors that we introduced in Eq. (20.51):

$$F^a{}_i = T^a_{ij}\phi_{0j}. \tag{21.36}$$

The matrix $F^a{}_i$ is not generally square; it has one row for each gauge generator, and one column for each component of ϕ. However, many of its elements are zero. Its nonzero elements connect the spontaneously broken gauge generators and the Goldstone bosons. In Eq. (20.56), we showed that the gauge boson masses generated through the Higgs mechanism can be written

$$m^2_{ab} = g^2 F^a{}_j F^b{}_j. \tag{21.37}$$

To give a concrete example of a matrix $F^a{}_j$, let us compute it in the GWS electroweak theory. Following the conventions introduced in Eq. (20.14), we should rewrite the Higgs field of the GWS model in terms of four real scalar fields. A convenient parametrization is

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} -i(\phi^1 - i\phi^2) \\ v + (h + i\phi^3) \end{pmatrix}. \tag{21.38}$$

The fields ϕ^i are the Goldstone bosons, and h is the massive Higgs boson. The vacuum state is simply

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}.$$

The real representation matrices are

$$T^a = -i\tau^a = -i\frac{\sigma^a}{2}, \qquad T^Y = -iY = -i\frac{1}{2}.$$

A simple computation then shows, for instance, that $T^1\phi_0$ equals $v/2$ times a unit vector in the ϕ^1 direction. Filling in the remaining components of $F^a{}_i$, with $a = 1, 2, 3, Y$ and $i = 1, 2, 3$, we find

$$gF^a{}_i = \frac{v}{2}\begin{pmatrix} g & 0 & 0 \\ 0 & g & 0 \\ 0 & 0 & g \\ 0 & 0 & -g' \end{pmatrix}. \tag{21.39}$$

We do not need to include the components of $F^a{}_i$ along the direction of the physical Higgs field h; the vectors $T^a\phi_0$ are all orthogonal to this direction.

If we insert (21.35) into (21.32) as a change of variables, we find, for the quadratic terms in the Lagrangian,

$$\begin{aligned} \mathcal{L}_2 = &-\tfrac{1}{2}A^a_\mu(-g^{\mu\nu}\partial^2 + \partial^\mu\partial^\nu)A^a_\nu + \tfrac{1}{2}(\partial_\mu\chi)^2 \\ &+ g\partial^\mu\chi_i A^a_\mu F^a{}_i + \tfrac{1}{2}(m_A^2)^{ab}A^a_\mu A^{\mu b} - \tfrac{1}{2}M_{ij}\chi_i\chi_j, \end{aligned} \tag{21.40}$$

where $(m_A^2)^{ab}$ is the gauge boson mass matrix (21.37) and

$$M_{ij} = \frac{\partial^2}{\partial\phi_i\partial\phi_j}V(\phi)\Big|_{\phi_0}. \tag{21.41}$$

We proved in Eq. (11.13) that

$$n_i M_{ij} = 0 \tag{21.42}$$

for all possible directions n_i in the subspace spanned by the $T^a\phi_0$, so the Goldstone bosons are massless.

To study the quantum theory of this system we start with the functional integral

$$Z = \int \mathcal{D}A\mathcal{D}\chi \, e^{i\int \mathcal{L}[A,\chi]}. \tag{21.43}$$

Using the Faddeev-Popov gauge-fixing procedure, we define this integral, analogously to (21.10), as

$$Z = C' \cdot \int \mathcal{D}A\mathcal{D}\chi \, \exp\Big[i\int d^4x\big(\mathcal{L}[A,\chi] - \tfrac{1}{2}(G)^2\big)\Big] \det\Big(\frac{\delta G}{\delta\alpha}\Big), \tag{21.44}$$

for an arbitrary gauge-fixing function $G(A,\chi)$. The R_ξ gauges are defined by the choice

$$G^a = \frac{1}{\sqrt{\xi}}\big(\partial_\mu A^{a\mu} - \xi g F^a{}_i \chi_i\big). \tag{21.45}$$

Note that G involves only the components of χ that lie in the subspace of the Goldstone bosons.

The gauge-fixing term adds to the Lagrangian the following set of quadratic terms:

$$(-\tfrac{1}{2}G^2)_2 = \tfrac{1}{2}A^a_\mu\big(\tfrac{1}{\xi}\partial^\mu\partial^\nu\big)A^a_\nu + g\partial_\mu A^{a\mu}F^a{}_i\chi_i - \tfrac{1}{2}\xi g^2\big[F^a{}_i\chi_i\big]^2. \tag{21.46}$$

The term that mixes A^a_μ and χ_i is arranged to cancel between (21.40) and (21.46). The final quadratic Lagrangian for the gauge and Goldstone boson fields is

$$\begin{aligned}\mathcal{L}_2 = -\frac{1}{2}A^a_\mu\Big(\Big[-g^{\mu\nu}\partial^2 + \big(1-\frac{1}{\xi}\big)\partial^\mu\partial^\nu\Big]\delta^{ab} - g^2 F^a{}_i F^b{}_i g^{\mu\nu}\Big)A^b_\nu \\ + \frac{1}{2}(\partial_\mu\chi)^2 - \frac{1}{2}\xi g^2 F^a{}_i F^a{}_j \chi_i\chi_j.\end{aligned} \tag{21.47}$$

The mass matrices of gauge bosons and Goldstone bosons in this Lagrangian are closely related to one another. The gauge boson mass matrix is

$$(m_A^2)^{ab} = g^2 F^a{}_i F^b{}_i = g^2(FF^T)^{ab}. \tag{21.48}$$

In an R_ξ gauge, the timelike components of the gauge bosons acquire the mass matrix

$$\xi m_A^2 = \xi g^2 (FF^T)^{ab}. \tag{21.49}$$

At the same time, the Goldstone bosons acquire the mass matrix

$$(m_G^2)_{ij} = \xi g^2 F^a{}_i F^a{}_j = \xi g^2 (F^T F)_{ij}. \tag{21.50}$$

The two matrices (21.49) and (21.50) have different numbers of zero eigenvalues, but their nonzero eigenvalues are in 1-to-1 correspondence. This is precisely the correspondence induced by the Higgs mechanism between the massive gauge bosons and the Goldstone bosons that they absorbed to gain mass.

Finally, we must construct the ghost Lagrangian. This is found from the gauge variation of the gauge-fixing term G^a. Inserting (21.30) and (21.31) into (21.45), we find

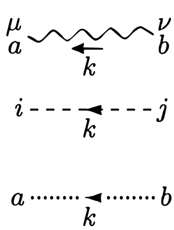
$$\frac{\delta G^a}{\delta \alpha^b} = \frac{1}{\sqrt{\xi}}\Big(\frac{1}{g}(\partial_\mu D^\mu)^{ab} + \xi g (T^a \phi_0)\cdot T^b(\phi_0 + \chi)\Big). \tag{21.51}$$

Thus, the ghost Lagrangian is

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a\big[-(\partial_\mu D^\mu)^{ab} - \xi g^2 (T^a\phi_0)\cdot T^b(\phi_0+\chi)\big]c^b. \tag{21.52}$$

Notice that the ghosts have exactly the same mass matrix (21.49) as the unphysical components of the gauge bosons. This Lagrangian also contains both the familiar coupling of the ghosts to the gauge fields and the coupling to the physical Higgs fields that we found in the Abelian case (21.15).

We have now computed the kinetic energy terms for gauge fields, scalar fields, and ghosts in an R_ξ gauge. It is straightforward to convert these results to the calculation of propagators for these fields; the computations are exactly the same as in the Abelian case. We find for the three propagators



$$= \left(\frac{-i}{k^2 - g^2FF^T}\left[g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2 - \xi g^2 FF^T}(1-\xi)\right]\right)^{ab},$$

$$= \left(\frac{i}{k^2 - \xi g^2 F^T F - M^2}\right)_{ij},$$

$$= \left(\frac{i}{k^2 - \xi g^2 FF^T}\right)^{ab}. \tag{21.53}$$

All of these equations involve the matrix F defined in Eq. (21.36); the appearance of a matrix in the denominator should be interpreted as a matrix inverse. The scalar field propagator also includes the mass matrix (21.41) of the physical Higgs bosons. There is no conflict between this matrix and the mass matrix of the Goldstone bosons, since they project onto orthogonal subspaces.

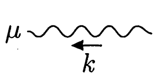
Although the preceding discussion has been extremely abstract, it is not hard to specialize to a particular example. So consider, once again, the GWS electroweak theory, for which the matrix $F^a{}_i$ is given by Eq. (21.39).

The gauge boson mass matrix in the GWS theory is

$$g^2FF^T = \frac{v^2}{4}\begin{pmatrix} g^2 & 0 & 0 & 0 \\ 0 & g^2 & 0 & 0 \\ 0 & 0 & g^2 & -gg' \\ 0 & 0 & -gg' & g'^2 \end{pmatrix},$$

in agreement with Eq. (20.124). (The g on the left-hand side should be interpreted as g' for the fourth component of F.) Diagonalizing this matrix gives the familiar relations (20.62). Thus, in the basis of mass eigenstates, the four

gauge-boson propagators decouple to give simply



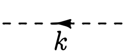
$$= \frac{-i}{k^2 - m^2}\left[g^{\mu\nu} - \frac{k^\mu k^\nu}{k^2 - \xi m^2}(1-\xi)\right], \qquad (21.54)$$

where m^2 is m_W^2, m_Z^2, or, for the photon, zero. Notice that, for the photon, this expression precisely reproduces Eq. (9.58).

The mass matrix of the Goldstone bosons in the GWS theory is

$$\xi g^2 F^T F = \xi \frac{v^2}{4} \begin{pmatrix} g^2 & 0 & 0 \\ 0 & g^2 & 0 \\ 0 & 0 & g^2 + g'^2 \end{pmatrix}.$$

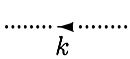
These fields therefore have the propagator



$$= \frac{i}{k^2 - \xi m^2}, \qquad (21.55)$$

with $m^2 = m_W^2$ for ϕ^1 and ϕ^2 (the bosons eaten by the $W^\pm$) and $m^2 = m_Z^2$ for ϕ^3 (the boson eaten by the Z). The field $h(x)$, which is the physical Higgs field, propagates independently with a mass determined by the Higgs potential (and no factor of ξ in the propagator).

Finally, there are four ghost fields. According to Eq. (21.53), these have the propagator



$$= \frac{i}{k^2 - \xi m^2}, \qquad (21.56)$$

with the same values of m^2 as the four gauge bosons.

The Feynman rules for the interaction vertices of these particles are complicated to write out, due to the large number of possible combinations. However, it is quite straightforward to generate these rules by expanding the weak interaction Lagrangian and reading off the vertices term by term. We will work out a few examples in the following section.*

21.2 The Goldstone Boson Equivalence Theorem

From the results of the previous section, we see that perturbative calculations in the R_ξ gauges involve intricate cancellations among unphysical particles. Sometimes, however, these unphysical particles can still leave their footprints in physical observables. In this section we will see that, in the high-energy limit, the unphysical Goldstone boson that is eaten by a massive gauge boson still controls the amplitude for emission or absorption of the gauge boson in its longitudinal polarization state.

*The complete Feynman rules for the weak-interaction gauge theory are given in Appendix B of Cheng and Li (1984).

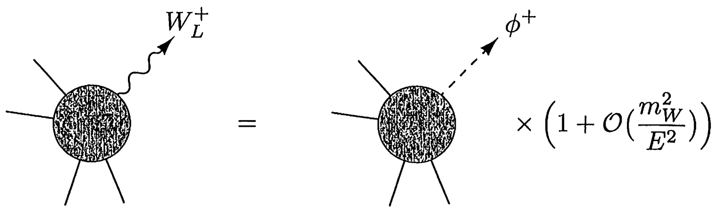


Figure 21.3. The Goldstone boson equivalence theorem. At high energy, the amplitude for emission or absorption of a longitudinally polarized massive gauge boson becomes equal to the amplitude for emission or absorption of the Goldstone boson that was eaten by the gauge boson.

When we introduced the Higgs mechanism for vector boson mass generation, we pointed out that it involves a certain conservation of degrees of freedom. A massless gauge boson, which has two transverse polarization states, combines with a scalar Goldstone boson to produce a massive vector particle, which has three polarization states. When the massive vector particle is at rest, its three polarization states are completely equivalent, but when it is moving relativistically, there is a clear distinction between the transverse and longitudinal polarization directions. This suggests that a rapidly moving, longitudinally polarized massive gauge boson might betray its origin as a Goldstone boson. The strongest version of this idea is expressed in Fig. 21.3: The amplitude for emission or absorption of a longitudinally polarized gauge boson becomes equal, at high energy, to the amplitude for emission or absorption of the Goldstone boson that was eaten. Remarkably, this statement is precisely correct, as a consequence of the underlying local gauge invariance. This *Goldstone boson equivalence theorem* was first proved by Cornwall, Levin, Tiktopoulos, and Vayonakis.†

Formal Aspects of Goldstone Boson Equivalence

The proof of the Goldstone boson equivalence theorem is based on the Ward identities of the spontaneously broken gauge theory. To give a complete proof of the theorem, we would have to construct and analyze these Ward identities in some detail. However, it is possible to understand the idea of the proof by examining the special case of the theorem in which a single massive vector boson is emitted or absorbed in a scattering process. The analysis of this special case requires only the relatively simple Ward identity satisfied by a current between on-shell states.‡

†J. M. Cornwall, D. N. Levin, and G. Tiktopoulos, *Phys. Rev.* **D10**, 1145 (1974); C. E. Vayonakis, *Lett. Nuov. Cim.* **17**, 383 (1976). For an illuminating discussion of the equivalence theorem, see B. W. Lee, C. Quigg, and H. Thacker, *Phys. Rev.* **D16**, 1519 (1977).

‡For a careful derivation of the equivalence theorem, including processes involving

To prepare for a discussion of longitudinal vector bosons, we need some simple kinematics. A vector boson at rest has momentum $k^\mu = (m, 0, 0, 0)$ and a polarization vector that is a linear combination of the three orthogonal unit vectors

$$(0, 1, 0, 0), \qquad (0, 0, 1, 0), \qquad (0, 0, 0, 1). \tag{21.57}$$

If we boost this particle along the $\hat{3}$ axis, its momentum boosts to $k^\mu = (E_{\mathbf{k}}, 0, 0, k)$. The three possible polarization vectors are now the three unit vectors satisfying

$$\epsilon^\mu k_\mu = 0, \qquad \epsilon^2 = -1. \tag{21.58}$$

Two of these are the first two vectors in (21.57); these give the transverse polarizations. The third vector satisfying (21.58) is the longitudinal polarization vector

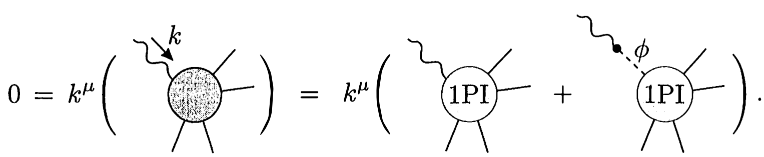
$$\epsilon_L^\mu(k) = \left(\frac{k}{m}, 0, 0, \frac{E_{\mathbf{k}}}{m}\right), \tag{21.59}$$

which is the boost of the third vector in (21.57). An important and somewhat counterintuitive feature of (21.59) is that it becomes increasingly parallel to k^μ as k becomes large. In fact, component by component,

$$\epsilon_L^\mu(k) = \frac{k^\mu}{m} + \mathcal{O}(m/E_{\mathbf{k}}) \tag{21.60}$$

as $k \to \infty$. Since the components of k^μ are growing as k, this statement is consistent with the requirement that $\epsilon_L \cdot k = 0$ while $k \cdot k = m^2$.

With this kinematic situation in mind, let us analyze the Ward identity satisfied by a gauge current matrix element between on-shell states. It is simplest to work in Lorentz gauge ($\xi = 0$), where the gauge-fixing term (21.45) does not involve the Goldstone boson fields. The Ward identity can then be written as follows:



(21.61)

In the last expression we have written the matrix element as the sum of two pieces. First, the current can couple directly into a one-particle-irreducible vertex function $\Gamma^\mu(k)$. This gives the class of diagrams that contribute to the scattering of a gauge boson from the external states. However, for a spontaneously broken gauge theory, there is an additional term, which is not one-particle-irreducible, in which the current creates a Goldstone boson and it is this particle that couples to the external states through a 1PI vertex $\Gamma(k)$.

Let us write the relation linking the gauge current and the Goldstone boson state as

$$\langle 0 | J^\mu | \pi(k) \rangle = -iFk^\mu, \tag{21.62}$$

multiple absorptions and emissions of massive vector bosons, see M. S. Chanowitz and M. K. Gaillard, *Nucl. Phys.* **B261**, 379 (1985).

as in Eq. (20.46). Then the argument leading to Eq. (20.56) tells us that the gauge boson mass is given by

$$m = gF, \tag{21.63}$$

where g is the gauge boson coupling constant.

With these identifications, we can write the Ward identity that follows from the conservation of the gauge current:

$$k_\mu \langle J^\mu \rangle = 0, \tag{21.64}$$

between on-shell states. Writing each term shown in (21.61) in terms of the appropriate one-particle-irredicible vertex function, we find

$$k_\mu \Gamma^\mu(k) + k_\mu \big(igFk^\mu\big)\frac{i}{k^2}\Gamma(k) = 0. \tag{21.65}$$

Thus,

$$k_\mu \Gamma^\mu(k) = m\Gamma(k). \tag{21.66}$$

Now use this equation in the limit of large gauge boson momentum. Since the gauge boson vertex is one-particle-irreducible, the momenta of propagators inside the vertex are not, in general, collinear with k^μ. Then, according to (21.60), we may replace k^μ/m by the longitudinal polarization vector. Notice that this would not be permissible (but, also, is not necessary) in the second term of (21.65). Our final result is

$$\epsilon_{L\mu}(k)\Gamma^\mu(k) = \Gamma(k), \tag{21.67}$$

as $k \to \infty$, with an error of order m^2/k^2. That is, in the high-energy limit, the couplings of longitudinal gauge bosons become precisely those of their associated Goldstone bosons.

The equivalence theorem can be derived in another way, using the counting of physical states in spontaneously broken gauge theories, which we discussed below Eq. (21.26). In the previous section, we saw that, at least at the tree level, unitarity is maintained in spontaneously broken gauge theories by the cancellation of diagrams that produce timelike-polarized gauge bosons against diagrams that produce Goldstone bosons.

The situation is most clear in Feynman-'t Hooft gauge. There, the numerator of the gauge boson propagator is $-g^{\mu\nu}$. We can write this in terms of polarization vectors as

$$-g^{\mu\nu} = \sum_{i=1,2,3} \epsilon_i^\mu(k)\epsilon_i^{\nu *}(k) - \frac{k^\mu k^\nu}{m^2}. \tag{21.68}$$

The last term is the contribution from unphysical timelike polarization states. The unitarity of the S-matrix requires that, when a Cutkosky cut through a diagram puts a gauge boson propagator on-shell, the contribution of this piece

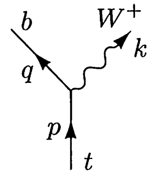
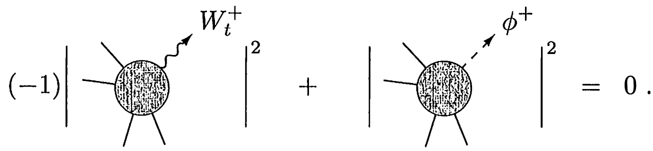


Figure 21.4. Decay of a t quark into $W^+ + b$.

must be canceled by a Cutkosky cut that runs through a Goldstone boson line. The required cancellation is

$$-\Big|\frac{k_\mu}{m}\Gamma^\mu(k)\Big|^2 + \big|\Gamma(k)\big|^2 = 0, \tag{21.69}$$

or, diagrammatically,



Once again, since $\Gamma^\mu(k)$ is a one-particle-irreducible vertex, we can use (21.60) to replace (k^μ/m) by the longitudinal polarization vector $\epsilon^\mu_L(k)$ for a high-energy gauge boson. Then (21.69) becomes just the square of (21.67).

Through these formal arguments, we can see, at least to the tree level in processes with single gauge boson emission, that the equivalence theorem must be valid. However, it is much more illuminating to see the equivalence theorem at work in explicit calculations for interesting physical processes. We will now illustrate its influence in two examples.

Top Quark Decay

The first example is the weak decay of the top quark. This charge $+2/3$ quark is sufficiently heavy that it can decay to a real W^+ through $t \to W^+ + b$. The diagram for this decay is given by the simple gauge vertex shown in Fig. 21.4.

Let us first try to guess the magnitude of the top quark width. The squared matrix element will contain a factor of g^2, times some expression with dimensions of mass. Since the width should be large if the top quark mass is heavy, a first guess might be

$$\Gamma \sim \frac{g^2}{4\pi} m_t. \tag{21.70}$$

The correct expression, however, turns out to be enhanced by a factor of $(m_t/m_W)^2$.

The amplitude for this decay can be read from Eq. (20.80):

$$i\mathcal{M} = \frac{ig}{\sqrt{2}}\bar{u}(q)\gamma^\mu\Big(\frac{1-\gamma^5}{2}\Big)u(p)\epsilon^*_\mu(k). \tag{21.71}$$

(We set the relevant CKM factor equal to 1.) We will now turn this amplitude into an expression for the decay rate of the top quark. For simplicity, we will ignore the mass of the b quark in this computation.

Squaring the amplitude in (21.71) according to our standard methods, and then averaging over initial and summing over final spins, we find

$$\tfrac{1}{2}\sum_{\text{spins}} |\mathcal{M}|^2 = \frac{g^2}{2}\left[q^\mu p^\nu + q^\nu p^\mu - g^{\mu\nu} q\cdot p\right] \sum_{\text{polarizations}} \epsilon^*_\mu(k)\epsilon_\nu(k). \tag{21.72}$$

We can sum explicitly over physical gauge boson polarizations by inserting the expression (21.26) for the polarization sum. This gives

$$\begin{aligned}\tfrac{1}{2}\sum_{\text{spins}} |\mathcal{M}|^2 &= \frac{g^2}{2}\left[q^\mu p^\nu + q^\nu p^\mu - g^{\mu\nu} q\cdot p\right]\left[-g_{\mu\nu} + \frac{k_\mu k_\nu}{m_W^2}\right]\\ &= \frac{g^2}{2}\left[q\cdot p + 2\frac{(k\cdot q)(k\cdot p)}{m_W^2}\right].\end{aligned} \tag{21.73}$$

For $m_b = 0$,

$$2q\cdot p = 2q\cdot k = m_t^2 - m_W^2, \qquad 2k\cdot p = m_t^2 + m_W^2. \tag{21.74}$$

Then

$$\tfrac{1}{2}\sum_{\text{spins}} |\mathcal{M}|^2 = \frac{g^2}{4}\frac{m_t^4}{m_W^2}\left(1 - \frac{m_W^2}{m_t^2}\right)\left(1 + 2\frac{m_W^2}{m_t^2}\right). \tag{21.75}$$

After multiplying by phase space, we find

$$\Gamma = \frac{g^2}{64\pi}\frac{m_t^3}{m_W^2}\left(1 - \frac{m_W^2}{m_t^2}\right)^2\left(1 + 2\frac{m_W^2}{m_t^2}\right). \tag{21.76}$$

This is larger than our initial estimate (21.70) by a factor $(m_t/m_W)^2$.

It is not difficult to find the origin of this enhancement, by using the Goldstone boson equivalence theorem. In the gauge theory of weak interactions, the top quark obtains its mass from its coupling to the Higgs sector. The relation between the top-Higgs coupling λ_t and the top quark mass is written in Eq. (20.103). The top quark can be heavy only if λ_t is large. But then the amplitude for the top quark to decay to a Goldstone boson will be enhanced above (21.70) by the factor

$$\frac{\lambda_t^2}{g^2} = \frac{m_t^2}{2m_W^2}, \tag{21.77}$$

which is in fact the enhancement we found in (21.76).

To make the comparison more precise, we will now compute the prediction of the equivalence theorem for the top quark decay rate into a longitudinally polarized W^+ boson. Recall from (20.101) that the term in the weak interaction Lagrangian that couples t and b to the Higgs field is

$$\Delta\mathcal{L} = -\lambda_t \epsilon^{ab}\overline{Q}_{La}\phi_b^\dagger t_R + \text{h.c.} \tag{21.78}$$

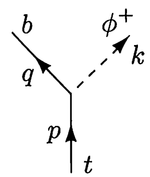


Figure 21.5. Decay of a t quark into a Goldstone boson and a b quark.

Decompose the Higgs field as in (21.38), and write

$$\phi^\pm = \frac{1}{\sqrt{2}}(\phi^1 \pm i\phi^2). \tag{21.79}$$

These are the fields of the charged Goldstone bosons that are eaten by the $W^\pm$. Including the Goldstone boson in the theory adds a process $t \to \phi^+ + b$, shown in Fig. 21.5. This process is mediated by the Lagrangian term

$$\Delta\mathcal{L} = \lambda_t \bar{b}_L \phi^+ t_R, \tag{21.80}$$

which leads to the decay amplitude

$$i\mathcal{M} = i\lambda_t \bar{u}(q)\Big(\frac{1+\gamma^5}{2}\Big)u(p). \tag{21.81}$$

From this expression, we easily find

$$\tfrac{1}{2}\sum_{\text{spins}} |\mathcal{M}|^2 = \lambda_t^2\, q\cdot p. \tag{21.82}$$

If we now ignore the mass of the Goldstone boson, or, equivalently, consider the limit $m_t \gg m_W$, we find for the top quark decay rate

$$\Gamma = \frac{\lambda_t^2}{32\pi}m_t = \frac{g^2}{64\pi}\frac{m_t^3}{m_W^2}, \tag{21.83}$$

in agreement with the leading term of (21.76) in this limit. Our results imply that only the production of the longitudinal polarization state of the W^+ is enhanced; this is easily checked directly by substituting explicit polarization vectors into (21.72).

In our derivation of (21.76), we summed over the physical polarization states of the emitted W^+; one might say that we used the prescription of the U gauge to sum over polarizations. We could equally well have used the prescription of Feynman-'t Hooft gauge, replacing

$$\sum_i \epsilon_\mu^*(k)\epsilon_\nu(k) \to -g_{\mu\nu}, \tag{21.84}$$

and also adding the contribution of the Goldstone boson emission diagram, treating the Goldstone boson as a massive particle with mass m_W. With these

prescriptions, the gauge boson matrix element gives

$$\frac{1}{2}\sum_{\text{spins}} |\mathcal{M}|^2 = \frac{g^2}{2}(2q\cdot p) = \frac{g^2}{2}(m_t^2 - m_W^2). \tag{21.85}$$

The Goldstone boson emission diagram gives

$$\frac{1}{2}\sum_{\text{spins}} |\mathcal{M}|^2 = \lambda_t^2\, q\cdot p = \frac{g^2}{4}\frac{m_t^2}{m_W^2}(m_t^2 - m_W^2). \tag{21.86}$$

The sum of these contributions indeed reproduces (21.75) and thus gives the same result (21.76) for the total decay rate. In Feynman-'t Hooft gauge, the enhancement due to the large coupling of the top quark to the Higgs sector shows up explicitly in the Goldstone boson emission contributions to the total rate of W^+ production.

$e^+e^- \to W^+W^-$

Our second example is more complicated, but also contains more interesting physics. This is the reaction $e^+e^- \to W^+W^-$. In this reaction, the equivalence theorem does not lead to an enhancement of the cross section, but, rather, directs a cancellation between Feynman diagrams. As we will see, this cancellation is essential for the internal consistency of the theory.

In Problem 9.1, we computed the cross section for e^+e^- annihilation into a pair of charged scalar particles, as in Fig. 21.6(a), and found the result

$$\frac{d\sigma}{d\cos\theta}(e^+e^- \to \phi^+\phi^-) = \frac{\pi\alpha^2}{4s}\sin^2\theta \tag{21.87}$$

at energies much larger than the scalar mass. Just as for e^+e^- annihilation to fermion pairs, this cross section falls as $1/s$ at high energy. It can be shown that this behavior is required by unitarity: Since the electron and positron annihilate through a pointlike vertex, the annihilation takes place in only one partial wave. Unitarity puts a limit on the amplitude in this partial wave, requiring that $\mathcal{M}$ be bounded by a constant, and thus that σ be bounded by $1/s$ at high energy.*

The same unitarity argument applies to e^+e^- annihilation to vector bosons. Here, however, it is much less obvious that Feynman diagrams actually produce a cross section consistent with unitarity. Consider the contribution of Fig. 21.6(b). We would expect that the square of this diagram should contain a contribution to the cross section of the form of the scalar contribution (21.87) multiplied by the dot product of polarization vectors:

$$\frac{d\sigma}{d\cos\theta}(e^+e^- \to W^+W^-) \sim \frac{\pi\alpha^2}{4s}\cdot \left|\epsilon(k_+)\cdot\epsilon(k_-)\right|^2, \tag{21.88}$$

*Partial-wave analysis for relativistic collisions is discussed in Perkins (1987), Chapter 4.

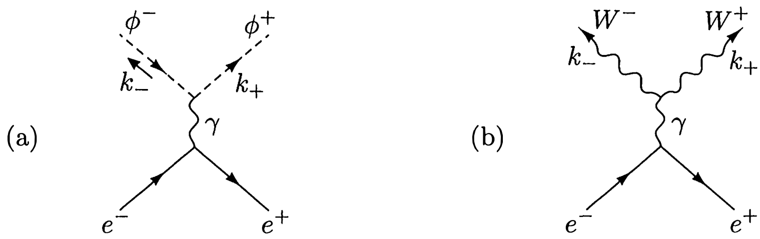


Figure 21.6. Electron-positron annihilation through a virtual photon (a) to charged scalar bosons, (b) to W bosons.

where k_+ and k_- are the momenta of the outgoing W bosons. For transversely polarized W bosons, this term is well behaved, but for longitudinally polarized W's it leads to problems. Using the approximation (21.60) for the longitudinal polarization vectors, we find

$$\epsilon(k_+) \cdot \epsilon(k_-) \rightarrow \frac{k_+ \cdot k_-}{m_W^2} \rightarrow \frac{s}{4m_W^2} \tag{21.89}$$

for $s \gg m_W^2$. This leads to a cross section that grows much faster than is allowed by unitarity. In principle, the cross section could be brought back down to a proper behavior by the addition of contributions from higher orders in perturbation theory, but this would be a most unpleasant resolution. It would imply that the theory of W bosons becomes strongly coupled at energies such that

$$\frac{s}{4m_W^2} \sim \left(\frac{g^2}{4\pi}\right)^{-1}, \tag{21.90}$$

corresponding to center-of-mass energies of order 1000 GeV. But if the theory of W bosons is strongly coupled at short distances, it is hard to understand why, at large distances, it should become the simple, weak-coupling theory that we observe.

Fortunately, there is another possible resolution of this problem. In the weak interaction gauge theory, there are three Feynman diagrams that contribute to the amplitude for $e^+e^- \rightarrow W^+W^-$ at the tree level; these are shown in Fig. 21.7. Each diagram separately produces a cross section that grows in the same manner as (21.88). However, it is possible that the badly behaved terms might cancel among the three diagrams, leaving a more proper high-energy behavior. If this miraculous cancellation were to occur, it would allow the theory of W bosons to be consistently weakly coupled up to very high energies.

Although such a cancellation seems unlikely at first sight, it is actually required by the Goldstone boson equivalence theorem. The theorem states that, at high energy, the cross section for producing longitudinal W bosons should be equal to the cross section for producing the corresponding scalar Goldstone bosons. But we know that scalar cross sections behave as $1/s$, as

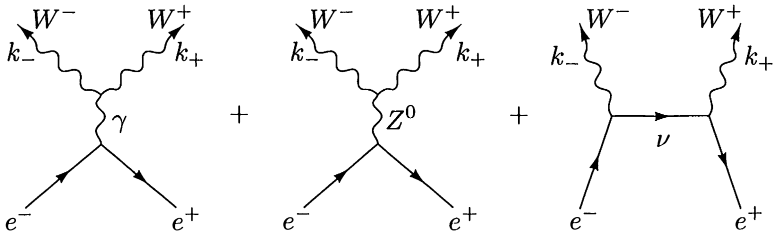


Figure 21.7. Diagrams contributing to $e^+e^- \to W^+W^-$ in the weak interaction gauge theory.

indicated in (21.87). Thus, somehow, the gauge boson cross section must also conspire to produce this result. We will now show this explicitly. We will see that the required cancellations are directed by the Ward identities of the gauge theory.

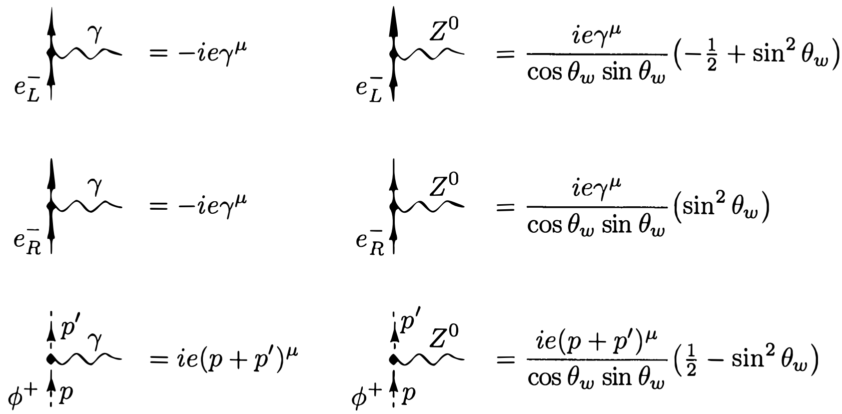
To prepare for this calculation, we need the Feynman rules for the vertices shown in Fig. 21.8. The Feynman rules for the couplings of the electron to W, Z, and γ can be read directly from (20.80). The relative strengths of these couplings are determined by the $SU(2) \times U(1)$ quantum numbers of the left- and right-handed components of the electron. It is equally straightforward to construct the couplings of the Goldstone bosons to Z and γ. Since the boson ϕ^+ has electric charge 1, the photon coupling is just that found in Problem 9.1. The Z coupling is determined with the additional information that the ϕ^+ has $I^3 = +1/2$. All of these expressions are shown in Fig. 21.8.

The three-gauge-boson vertices that appear in Fig. 21.7 arise from the cubic terms in the gauge field action. Since the $U(1)$ field strength is linear in gauge fields, these come only from the kinetic term of the $SU(2)$ gauge field. To identify the specific pieces we need, we must rewrite this cubic term in the basis of mass eigenstates given by (20.63) and (20.64). This can be done as follows:

$$\begin{aligned}
-\tfrac{1}{4}(F^a_{\mu\nu})^2 &\to -\tfrac{1}{2}(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu) g\epsilon^{abc} A^{\mu b} A^{\nu c} \\
&= -g(\partial_\mu A^1_\nu - \partial_\nu A^1_\mu) A^{\mu 2} A^{\nu 3} + g(\partial_\mu A^2_\nu - \partial_\nu A^2_\mu) A^{\mu 1} A^{\nu 3} \\
&\quad - g(\partial_\mu A^3_\nu - \partial_\nu A^3_\mu) A^{\mu 1} A^{\nu 2} \\
&= ig\big[(\partial_\mu W^+_\nu - \partial_\nu W^+_\mu) W^{\mu -} A^{\nu 3} - (\partial_\mu W^-_\nu - \partial_\nu W^-_\mu) W^{\mu +} A^{\nu 3} \\
&\quad + \tfrac{1}{2}(\partial_\mu A^3_\nu - \partial_\nu A^3_\mu)(W^{\mu +} W^{\nu -} - W^{\mu -} W^{\nu +})\big]. \qquad (21.91)
\end{aligned}$$

Finally, inserting $A^3_\mu = \cos\theta_w Z_\mu + \sin\theta_w A_\mu$ and $g = e/\sin\theta_w$, we find the Feynman rules shown in Fig. 21.9.

Before examining the amplitude for e^+e^- annihilation to vector boson

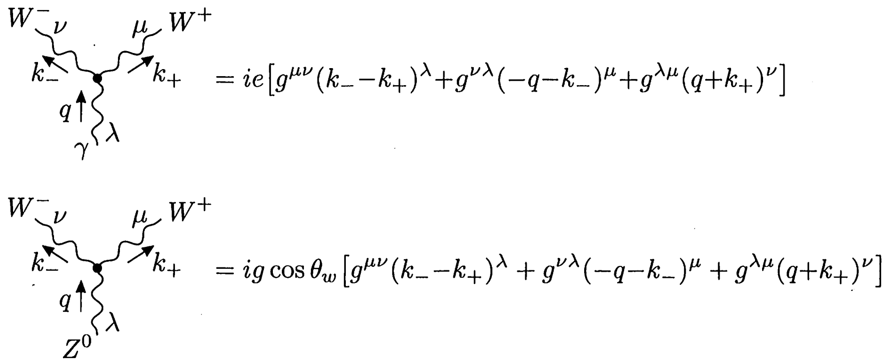


$$= -ie\gamma^\mu \qquad = \frac{ie\gamma^\mu}{\cos\theta_w \sin\theta_w}\left(-\tfrac{1}{2} + \sin^2\theta_w\right)$$

$$= -ie\gamma^\mu \qquad = \frac{ie\gamma^\mu}{\cos\theta_w \sin\theta_w}\left(\sin^2\theta_w\right)$$

$$= ie(p+p')^\mu \qquad = \frac{ie(p+p')^\mu}{\cos\theta_w \sin\theta_w}\left(\tfrac{1}{2} - \sin^2\theta_w\right)$$

Figure 21.8. Feynman rules of the weak-interaction gauge theory for electrons and scalars coupling to photons and Z bosons.



$$= ie\big[g^{\mu\nu}(k_- - k_+)^\lambda + g^{\nu\lambda}(-q-k_-)^\mu + g^{\lambda\mu}(q+k_+)^\nu\big]$$

$$= ig\cos\theta_w\big[g^{\mu\nu}(k_- - k_+)^\lambda + g^{\nu\lambda}(-q-k_-)^\mu + g^{\lambda\mu}(q+k_+)^\nu\big]$$

Figure 21.9. Feynman rules of the weak-interaction gauge theory for $WW\gamma$ and WWZ vertices.

pairs, we will first work out the amplitude for production of a pair of charged scalars. The equivalence theorem predicts that the amplitude for production of two longitudinal W bosons should become equal to this amplitude at high energy. Assembling vertices from Fig. 21.8, we find that, for an electron of either helicity, the amplitude to annihilate to scalars through a virtual photon is

$$i\mathcal{M}(ee \to \gamma^* \to \phi^+\phi^-) = ie^2\bar{v}\gamma_\mu u \frac{1}{q^2}(k_+ - k_-)^\mu, \tag{21.92}$$

where k_+, k_- are the momenta of the scalars and $q = k_+ + k_-$. The corresponding amplitude for annihilation through a virtual Z^0 depends on the e^+e^- helicities. Adding these contributions to the preceding expression, we

find

$$i\mathcal{M}(e^-_L e^+_R \to \phi^+\phi^-) = ie^2 \bar{v}_L \gamma_\mu u_L \left[\frac{1}{q^2} + \frac{(\frac{1}{2} - \sin^2\theta_w)^2}{\sin^2\theta_w \cos^2\theta_w} \frac{1}{q^2 - m_Z^2}\right](k_+ - k_-)^\mu,$$
$$i\mathcal{M}(e^-_R e^+_L \to \phi^+\phi^-) = ie^2 \bar{v}_R \gamma_\mu u_R \left[\frac{1}{q^2} - \frac{(\frac{1}{2} - \sin^2\theta_w)}{\cos^2\theta_w} \frac{1}{q^2 - m_Z^2}\right](k_+ - k_-)^\mu. \tag{21.93}$$

Notice that, in the high-energy limit, the amplitude for the annihilation of right-handed electrons cancels down to

$$i\mathcal{M}(e^-_R e^+_L \to \phi^+\phi^-) \to i\frac{e^2}{2\cos^2\theta_w} \bar{v}_R \gamma_\mu u_R \frac{1}{q^2}(k_+ - k_-)^\mu, \tag{21.94}$$

which is just the amplitude for an e^-_R, with $Y = -1$, to couple to a ϕ^+, with $Y = 1/2$, through the $U(1)$ gauge boson B_μ with coupling constant $g' = e/\cos\theta_w$. This expression reflects the fact that the e^-_R has no direct coupling to the $SU(2)$ gauge bosons. Similarly, the amplitude for left-handed electrons tends to

$$i\mathcal{M}(e^-_L e^+_R \to \phi^+\phi^-) \to ie^2 \left[\frac{1}{4\cos^2\theta_w} + \frac{1}{4\sin^2\theta_w}\right] \bar{v}_L \gamma_\mu u_L \frac{1}{q^2}(k_+ - k_-)^\mu \tag{21.95}$$

in the high-energy limit. This has the structure of a coherent sum of amplitudes with B_μ and A^3_μ exchange. In just the way that we saw in Chapter 11, the symmetry structure of a gauge theory with spontaneously broken symmetry is recovered in the high-energy limit.

Now let us compare these results to a direct calculation of the W^+W^- production amplitude in the weak interaction gauge theory. Begin with the case of an initial e^-_R. Since the coupling of the electron to the W^- is purely left-handed, the third diagram of Fig. 21.7 vanishes in this case, so the computation is a bit easier. The first two diagrams of Fig. 21.7 have exactly the same structure and sum to

$$i\mathcal{M}(e^-_R e^+_L \to W^+W^-) = \bar{v}_R \gamma_\lambda u_R \left[(-ie)\frac{-i}{q^2}(ie) + \frac{ie\sin\theta_w}{\cos\theta_w} \frac{-i}{q^2 - m_Z^2} \frac{ie\cos\theta_w}{\sin\theta_w}\right]$$
$$\cdot \left[g^{\mu\nu}(k_- - k_+)^\lambda + g^{\lambda\nu}(-q - k_-)^\mu + g^{\lambda\mu}(k_+ + q)^\nu\right]\epsilon^*_\mu(k_+)\epsilon^*_\nu(k_-). \tag{21.96}$$

This equation is valid in any of the R_ξ gauges, since, if we ignore the electron mass,

$$q^\lambda \bar{v}_R \gamma_\lambda u_R = 0. \tag{21.97}$$

The second line of Eq. (21.96) contains the enhancement for longitudinal W bosons mentioned above. If we approximate the longitudinal polarization vectors by (21.60) and drop terms that do not grow as $s \to \infty$, this line becomes

$$\left[g^{\mu\nu}(k_- - k_+)^\lambda + g^{\lambda\nu}(-q - k_-)^\mu + g^{\lambda\mu}(k_+ + q)^\nu\right]\frac{k_{+\mu}}{m_W}\frac{k_{-\nu}}{m_W}$$

$$
\begin{aligned}
&= \frac{1}{m_W^2}\left[k_+\cdot k_-(k_- - k_+)^\lambda - 2k_-\cdot k_+ k_-^\lambda + 2k_+\cdot k_- k_+^\lambda\right] + \mathcal{O}(1)\cdot(k_- - k_+)^\lambda \\
&= \frac{s}{2m_W^2}(k_+ - k_-)^\lambda + \cdots . \qquad (21.98)
\end{aligned}
$$

On the other hand, the expression in brackets in the first line of (21.96) cancels almost completely, to

$$
-ie^2\left(\frac{1}{q^2} - \frac{1}{q^2 - m_Z^2}\right) = +ie^2\frac{m_Z^2}{q^2(q^2 - m_Z^2)}.
$$

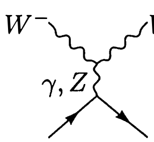
Using both of these simplifications, we find

$$
i\mathcal{M}(e_R^- e_L^+ \to W_L^+ W_L^-) = \bar{v}_R\gamma_\lambda u_R\left[(ie^2)\frac{m_Z^2}{s^2}\right]\frac{s}{2m_W^2}(k_+ - k_-)^\lambda. \qquad (21.99)
$$

By inserting the relation $m_W = m_Z\cos\theta_w$, we see that this amplitude is identical to (21.94), as required by the equivalence theorem.

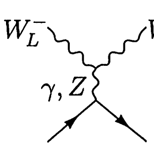
For the amplitude with an initial e_L^-, the computation is somewhat more involved. Now all three diagrams of Fig. 21.7 contribute, and since the last diagram has a different kinematic structure, it will be less clear how the diagrams combine together. In what follows, we will demonstrate the cancellation of the unitarity-violating enhanced terms, and we will indicate how the terms one order smaller in m_W^2/s assemble into the correct structure. However, we will not account rigorously for all of these smaller terms. The full calculation of these diagrams is the subject of Problem 21.2.

For the case of an initial e_L^-, the first two diagrams of Fig. 21.7 sum to the expression



$$
\begin{aligned}
&= \bar{v}_L\gamma_\lambda u_L\left[(-ie)\frac{-i}{q^2}(ie) + \frac{ie(-\frac{1}{2}+\sin^2\theta_w)}{\sin\theta_w\cos\theta_w}\frac{-i}{q^2 - m_Z^2}\frac{ie\cos\theta_w}{\sin\theta_w}\right] \\
&\quad\cdot\left[g^{\mu\nu}(k_- - k_+)^\lambda + g^{\lambda\nu}(-q - k_-)^\mu + g^{\lambda\mu}(k_+ + q)^\nu\right]\epsilon_\mu^*(k_+)\epsilon_\nu^*(k_-), \qquad (21.100)
\end{aligned}
$$

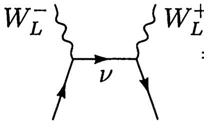
which differs from (21.96) only in the coupling of the electron to the virtual Z^0. For longitudinal W bosons, we can simplify this expression as we did (21.96), obtaining



$$
= \bar{v}_L\gamma_\lambda u_L(ie^2)\left[\frac{m_Z^2}{s(s - m_Z^2)} - \frac{1}{2\sin^2\theta_w}\frac{1}{s - m_Z^2}\right]\frac{s}{2m_W^2}(k_+ - k_-)^\lambda. \qquad (21.101)
$$

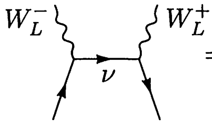
The second term in brackets is a potentially dangerous contribution, which must be canceled by the diagram with t-channel neutrino exchange. This

diagram has the value



$$= \left(\frac{ig}{\sqrt{2}}\right)^2 \bar{v}_L \gamma^\mu \frac{i(\not{\ell} - \not{k}_-)}{(\ell - k_-)^2} \gamma^\nu u_L(\ell) \epsilon^*_\mu(k_+) \epsilon^*_\nu(k_-), \tag{21.102}$$

where ℓ is the initial electron momentum. Approximating the longitudinal polarization vectors as before, we have

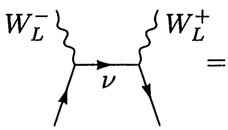


$$= -i\frac{g^2}{2} \bar{v}_L \frac{\not{k}_+}{m_W} \frac{(\not{\ell} - \not{k}_-)}{(\ell - k_-)^2} \frac{\not{k}_-}{m_W} u_L(\ell). \tag{21.103}$$

Now we manipulate this expression as if we were proving a Ward identity. Using the fact that $u_L(\ell)$ satisfies the Dirac equation,

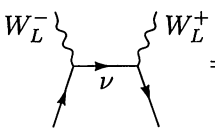
$$(\not{\ell} - \not{k}_-)\not{k}_- u_L(\ell) = -(\not{\ell} - \not{k}_-)^2 u_L(\ell) = -(\ell - k_-)^2 u_L(\ell), \tag{21.104}$$

expression (21.103) reduces to



$$= i\frac{g^2}{2} \bar{v}_L \frac{\not{k}_+}{m_W^2} u_L(\ell). \tag{21.105}$$

Finally, using Eq. (21.97), we can rewrite this expression as



$$= ie^2 \frac{1}{2\sin^2\theta_w} \frac{1}{2m_W^2} \bar{v}_L \gamma_\lambda u_L(\ell)(k_+ - k_-)^\lambda. \tag{21.106}$$

This term cancels the dangerous high-energy behavior of (21.101). To see that the sum of diagrams has the correct high-energy limit, however, the approximations that we have used are not quite adequate. In particular, the correction to relation (21.60) for the polarization vectors is of order m_W^2/s and must be taken into account. When all of the corrections of order m_W^2/s are included, it turns out that the sum of the s-channel diagrams (21.101) is unchanged, while the expression for the neutrino exchange diagram (21.106) is multiplied by the factor $(1 + 2m_W^2/s)$. Then the sum of all three diagrams gives

$$\begin{aligned} i\mathcal{M}(e_L^- e_R^+ \to W_L^+ W_L^-) &= ie^2 \bar{v}_L \gamma_\lambda u_L (k_+ - k_-)^\lambda \frac{1}{s} \\ &\cdot \left[\frac{1}{2\cos^2\theta_w} - \frac{1}{4\cos^2\theta_w \sin^2\theta_w} + \frac{1}{2\sin^2\theta_w}\right]. \end{aligned} \tag{21.107}$$

The middle term in brackets cancels half of each of the other two terms, to give an expression that agrees precisely with Eq. (21.95).

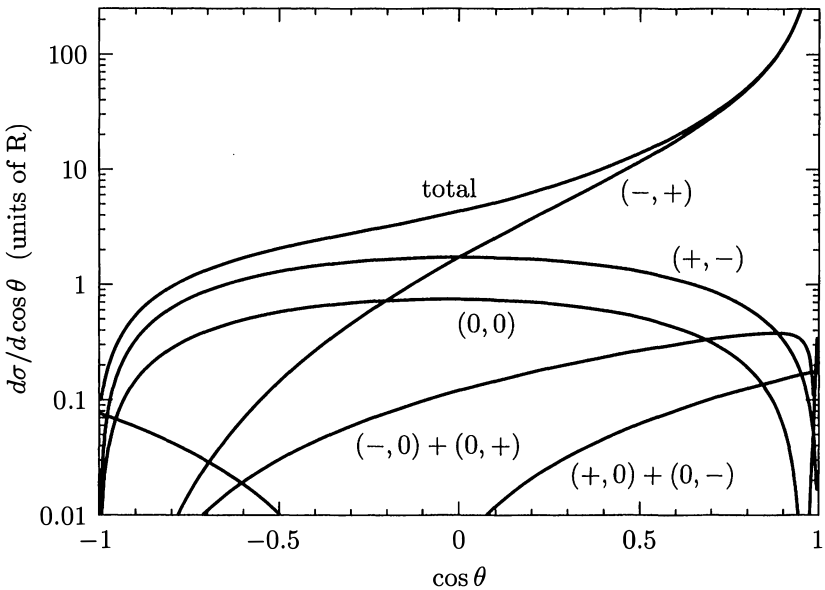


Figure 21.10. The differential cross section for $e^-_L e^+_R \to W^+W^-$, in units of R (Eq. (5.15)), at $E_{\rm cm} = 1000$ GeV. The various curves show the contributions to the total from individual helicity states of W^- and W^+; these are denoted (h_-, h_+), where each helicity takes the values $(+,-,0)$. The contributions from the $(+,+)$ and $(-,-)$ states are too small to be visible. Notice that both the $W^-_L W^+_L$ cross section, denoted $(0,0)$, and the $(+,-)$ cross section become proportional to $\sin^2\theta$ at very high energy.

The calculation of Problem 21.2 gives for the complete annihilation amplitude

$$\begin{aligned} i\mathcal{M}(e^-_L e^+_R \to W^+_L W^-_L) &= ie^2 \bar{v}_L \gamma_\lambda u_L (k_+ - k_-)^\lambda \frac{1}{s} \\ &\cdot \left[\frac{1}{2\sin^2\theta_w} \left\{ -\frac{s}{s-m_Z^2}\left(\frac{m_Z^2}{2m_W^2}+1\right) + \frac{2}{\beta^2} \right.\right. \\ &\quad \left.\left. - \frac{8m_W^2}{s\beta^2(1+\beta^2-2\beta\cos\theta)} \right\} + \frac{m_Z^2}{m_W^2}\left(\frac{\frac{1}{2}s+m_W^2}{s-m_Z^2}\right)\right], \end{aligned} \tag{21.108}$$

where $\beta = (1 - 4m_W^2/s)^{1/2}$ is the W boson velocity. The high-energy limit of this expression indeed reproduces (21.107). The contributions to the differential cross section for $e^-_L e^+_R \to W^+W^-$ from this and the other possible helicity states are plotted in Fig. 21.10.

These cancellations among the diagrams of Fig. 21.7 occur by virtue of the Ward identities of the gauge theory. That is, they occur only because the theory has an underlying local gauge invariance. At the beginning of our discussion, we argued that these cancellations are necessary to insure that the theory remains, in a consistent way, weakly coupled up to arbitrarily

high energy. In Section 20.1, we showed that one can generate masses for vector bosons by spontaneously breaking local gauge invariance. We have now argued the converse of that result: that the only theories of massive vector bosons that do not have violent high-energy behavior are those that result from spontaneously broken gauge theories.†

21.3 One-Loop Corrections to the Weak-Interaction Gauge Theory

The final topic in our study of spontaneously broken gauge theories is the computation of one-loop corrections in the weak-interaction gauge theory. As we discussed in Section 20.2, tree-level diagrams produce a number of intricate predictions for the couplings of the Z^0 and the cross sections for neutral current reactions. In general, these predictions are modified by the effects of one-loop diagrams. In this section we will study some examples of these one-loop corrections.

As in any renormalizable field theory, the one-loop diagrams of the electroweak gauge theory are typically ultraviolet divergent. These divergences can be absorbed by adjusting the underlying parameters of the theory. These adjustments define a set of counterterms which, by renormalizability, render the full set of one-loop diagrams of the theory finite. Those amplitudes that are not adjusted by hand then become predictions of the theory.

In Chapter 11, we saw that this general procedure, which applies to any renormalizable field theory, gives especially rich information when applied to a theory with spontaneous symmetry breaking. In a theory with spontaneously broken symmetry, the amplitudes of the theory vary markedly for different particles in the same multiplet of the original symmetry. However, the counterterms of the theory respect the symmetry relations. Thus, the adjustment of an amplitude for one particle leads to definite predictions for other particles that are not related by any manifest symmetry.

Theoretical Orientation, and a Specific Problem

At the end of Section 11.6, we presented a useful framework for organizing calculations of the predictions of renormalizable theories with spontaneous symmetry breaking. We defined a *zeroth-order natural relation* to be a relation among observable quantities in the theory that is true for any values of the parameters in the Lagrangian. Since the counterterms of the theory shift the values of the underlying parameters without adding new terms, a zeroth-order natural relation will not be corrected by these counterterms. Thus, if the theory is renormalizable, the one-loop corrections to a zeroth-order natural

†This statement is proved systematically in the paper of Cornwall, Levin, and Tiktopoulos cited at the beginning of this section.

relation will be finite, and will in fact be definite predictions from the quantum structure of the field theory. Though we discussed this idea originally in theories with spontaneously broken global symmetry, it applies equally well to theories with spontaneously broken gauge symmetry. In this section, we will apply this idea to derive finite one-loop corrections to relations in the weak-interaction gauge theory.

It is easy to find zeroth-order natural relations in the electroweak theory. The leading-order predictions given in Section 20.2 involve a relatively small number of free parameters. Many of these predictions are made for energies at which the quark and lepton masses can be ignored; then they depend only on the coupling constants g and g' and the vacuum expectation value v, which sets the scale of spontaneous symmetry breaking. The remaining ingredients of the weak-interaction theory are given in terms of these parameters; for example,

$$\begin{aligned} m_W &= g\frac{v}{2}, & m_Z &= \sqrt{g^2+g'^2}\frac{v}{2}, \\ e &= \frac{gg'}{\sqrt{g^2+g'^2}}, & \frac{G_F}{\sqrt{2}} &= \frac{g^2}{8m_W^2} = \frac{1}{2v^2}. \end{aligned} \tag{21.109}$$

Even in this set of quantities, we have four relations that depend on three underlying parameters, so there is one relation of observable quantities that is independent of the parameters of the Lagrangian.

Since many of the predictions of the weak interaction gauge theory are determined by the parameter $\sin^2\theta_w$, it is useful to define $\sin^2\theta_w$ in terms of observables and then use this definition as a basis for constructing natural relations. In our discussion of the precision tests of electroweak theory in Section 20.2, we used the definition

$$s_W^2 \equiv 1 - \frac{m_W^2}{m_Z^2} \tag{21.110}$$

as a standard for comparison of different experiments. But since the three most accurately known weak-interaction observables are α, G_F, and m_Z, it is useful to construct another physical definition of $\sin^2\theta_w$ based on these three quantities. Define θ_0 such that

$$\sin 2\theta_0 \equiv \left(\frac{4\pi\alpha_*}{\sqrt{2}G_F m_Z^2}\right)^{1/2}, \tag{21.111}$$

where α_* is the running coupling constant of QED evaluated at the scale $Q^2 = m_Z^2$. The renormalization group insists that it is the value of the electric charge at the weak-interaction scale that enters precision electroweak predictions, and this observation is confirmed by summing radiative correction diagrams involving light quarks and leptons. The current best values of the quantities in Eq. (21.111) give

$$s_0^2 \equiv \sin^2\theta_0 = 0.2307 \pm 0.0005. \tag{21.112}$$

Thus, this quantity provides a very accurate standard of reference.

Once Eq. (21.111) is taken to define a reference value of $\sin^2\theta_w$, the equations of Section 20.2 that connect $\sin^2\theta_w$ to other observables become zeroth-order natural relations. For example, the tree-level equations

$$\frac{m_W^2}{m_Z^2} = \cos^2\theta_w, \qquad A_{LR}^e = \frac{(\frac{1}{2} - \sin^2\theta_w)^2 - (\sin^2\theta_w)^2}{(\frac{1}{2} - \sin^2\theta_w)^2 + (\sin^2\theta_w)^2} \tag{21.113}$$

are natural relations linking four observables of the weak interactions. The corrections to these relations will be well-defined predictions of the theory.

In principle, we could now compute all of the one-loop diagrams that correct the parameters m_W, m_Z, G_F, α, and A_{LR}^e. However, this is a very complicated exercise, requiring an extensive technical apparatus.‡ In this section we will focus on radiative corrections from one simple source that can be considered independently. Aside from the question of anomalies, the electroweak theory does not restrict the number of quark or lepton generations. Thus, it is sensible, and gauge invariant, to compute the one-loop corrections due to one quark or lepton doublet. For definiteness, we consider the effects of the (t, b) quark doublet.

By focusing on the radiative corrections due to heavy quarks, we dramatically simplify the calculational task before us. The various observables of the weak-interaction gauge theory are extracted from the measurement of scattering amplitudes with light fermions, leptons or quarks, in the initial and final states. For example, G_F is measured from the strength of a low-energy weak-interaction process, usually chosen to be the rate of muon decay: $\mu \to \nu_\mu e^- \bar{\nu}_e$. For any such process, there are one-loop corrections of many kinds, as shown in Fig. 21.11. In addition to corrections to the vector boson propagator, there are vertex corrections, box diagrams, and diagrams with real photon emission. In general, the contributions of the various classes of diagrams are not gauge invariant; rather, gauge invariance results from cancellations between the classes of diagrams in Fig. 21.11(b), (c), and (d). However, since heavy quarks do not couple directly to the light leptons, the (t, b) doublet contributes only the single diagram shown in Fig. 21.11(f), which must be gauge invariant by itself. This same conclusion applies to the (t, b) correction to other leptonic weak interaction processes. If we ignore the CKM angles that mix the t and b with other species, the conclusion extends also to weak-interaction processes involving light quarks.

A similar situation occurs with other species of particles, such as those of the Higgs sector. The coupling of Higgs sector particles to a light quark or lepton is proportional to the fermion's mass, which we can often ignore. Thus the most important contributions from Higgs-sector particles are propagator corrections. The case in which the spontaneous symmetry breaking is produced by a single scalar field ϕ is particularly straightforward to analyze;

‡A detailed theoretical discussion of one-loop corrections to the electroweak theory can be found in W. Hollik, *Fortscr. d. Physik* **38**, 165 (1990).

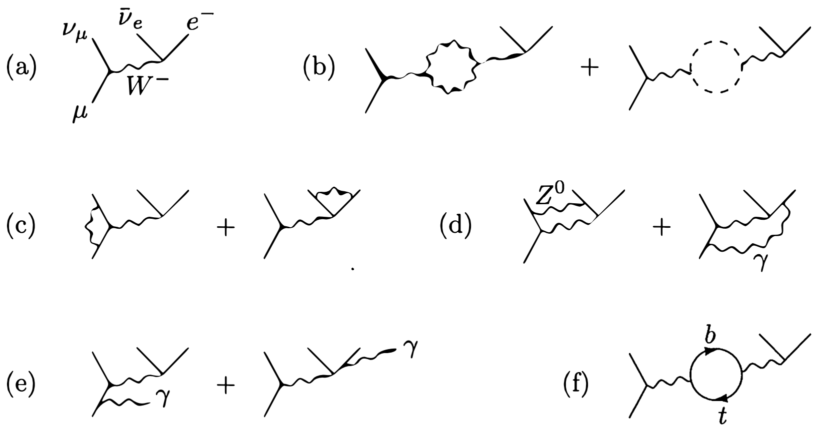


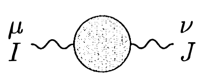
Figure 21.11. Examples of radiative corrections to μ decay in the weak interaction gauge theory: (a) lowest-order diagram; (b) propagator corrections; (c) vertex diagrams; (d) box diagrams; (e) real photon corrections; (f) the contribution of the (t, b) doublet.

this is done in Problem 21.4. Loop corrections from particles that do not couple directly to the external fermions are often termed *oblique*, since they enter the low-energy weak interactions only indirectly.

Influence of Heavy Quark Corrections

Our task, then, is to compute the corrections to relations (21.113) due to the (t, b) doublet. These two relations depend on five observable quantities—m_Z, m_W, A^e_{LR}, α, and G_F—with the last two parameters entering through θ_w and Eq. (21.111). We will express these five quantities as functions of the bare parameters g, g', and v, with corrections proportional to combinations of t and b vacuum polarization diagrams. The zeroth-order terms will naturally cancel out when we compute the corrections to the relations (21.113).

The loop amplitudes that we require are shown in Fig. 21.12. To deal with these contributions most straightforwardly, we introduce a uniform notation for vacuum polarization amplitudes. Denote the vacuum polarization amplitude involving the gauge bosons I and J as



$$= i\Pi^{\mu\nu}_{IJ}(q), \tag{21.114}$$

where I and J may be γ, W, or Z. When the gauge bosons are massive, the vacuum polarization amplitudes need not be transverse by themselves, so $\Pi^{\mu\nu}_{IJ}(q)$ need not vanish at $q^2 = 0$. Thus, we will change our notation from the case of QED and write the decomposition of $\Pi^{\mu\nu}_{IJ}(q)$ into tensor structures as

$$\Pi^{\mu\nu}_{IJ}(q) = \Pi_{IJ}(q^2)g^{\mu\nu} - \Delta(q^2)q^\mu q^\nu. \tag{21.115}$$

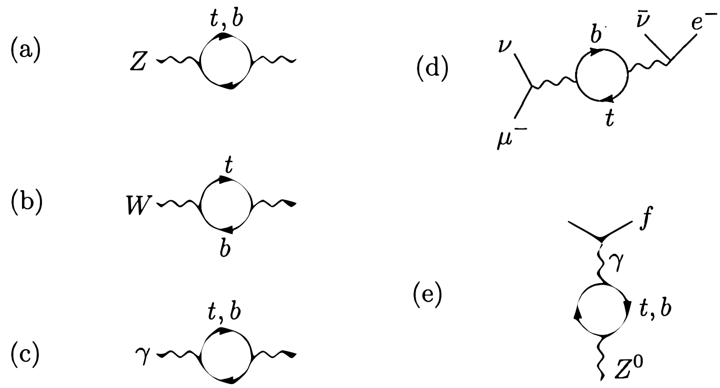


Figure 21.12. One-loop corrections from t and b to weak-interaction observables: (a) m_Z; (b) m_W; (c) α; (d) G_F; (e) A^e_{LR}.

In all of the examples to follow, the factors q^μ will dot into currents of light leptons, to give zero as in Eq. (21.97). Thus the form factor $\Delta(q^2)$ will drop out of our calculations. Our previous result that $\Pi^{\mu\nu}(q)$ vanishes in QED at $q^2 = 0$ appears in this formalism as the set of constraints

$$\Pi_{\gamma\gamma}(0) = \Pi_{\gamma Z}(0) = 0. \tag{21.116}$$

For the other amplitudes, our sign conventions are chosen so that a positive value of $\Pi_{IJ}(m^2)$ gives a positive mass shift to the gauge boson. Let us also define

$$\Pi'_{\gamma\gamma}(0) = \left.\frac{d\Pi_{\gamma\gamma}}{dq^2}\right|_{q^2=0}; \tag{21.117}$$

this is the quantity we called $\Pi(0)$ in Eq. (7.73).

Now we use this notation to write the loop corrections to each of the five observables. The first two diagrams in Fig. 21.12 are simply mass corrections, and so, straightforwardly,

$$\begin{aligned} m_Z^2 &= (g^2 + g'^2)\frac{v^2}{4} + \Pi_{ZZ}(m_Z^2), \\ m_W^2 &= g^2\frac{v^2}{4} + \Pi_{WW}(m_W^2). \end{aligned} \tag{21.118}$$

Note that both vacuum polarization amplitudes are evaluated at the poles in the respective propagators. To evaluate the shift of α by one-loop corrections, we consider the effect of Fig. 21.12(c) on the low-energy Coulomb potential. The values of the leading-order propagator and the one-loop correction combine to give the factors

$$\frac{-ie^2}{q^2}\left(1 + i\Pi_{\gamma\gamma}(q^2)\cdot\frac{-i}{q^2}\right), \tag{21.119}$$

where, in this equation, e^2 is given in terms of bare variables as in (21.109). Thus, the observed value of α, in the limit $q^2 \to 0$, is modified according to the relation

$$4\pi\alpha = \frac{g^2 g'^2}{g^2 + g'^2}\big(1 + \Pi'_{\gamma\gamma}(0)\big). \tag{21.120}$$

In a similar way, the diagrams of Fig. 21.12(d) give a modified strength of the 4-fermion weak interaction process that leads to μ decay. The leading and one-loop diagrams sum to

$$\frac{g^2}{q^2 - m_W^2}\Big(1 + i\Pi_{WW}(q^2)\frac{-i}{q^2 - m_W^2}\Big). \tag{21.121}$$

Then the effective strength of the weak interaction vertex at $q^2 = 0$ is shifted as follows:

$$\frac{G_F}{\sqrt{2}} = \frac{1}{2v^2}\Big(1 - \frac{\Pi_{WW}(0)}{m_W^2}\Big). \tag{21.122}$$

Notice that, in the approximation of keeping only oblique corrections, the strength of every low-energy weak interaction amplitude is corrected by this same factor.

Finally, the polarization asymmetry A_{LR}^e is corrected by a (t, b) loop diagram according to Fig. 21.12(e). The analogous diagram with an intermediate Z^0 is summed into the Z^0 propagator and does not affect the form of the vertex. At zeroth order, the coupling of the Z^0 to any left- or right-handed light fermion is given, according to Eq. (20.71), by

$$\sqrt{g^2 + g'^2}\Big(T^3 - \frac{g'^2}{g^2 + g'^2}Q\Big). \tag{21.123}$$

The coefficient of Q is the bare value of $\sin^2\theta_w$. The loop diagram in Fig. 21.12(e) adds to this a contribution

$$i\Pi_{Z\gamma}(q^2)\frac{-i}{q^2}\cdot(ieQ). \tag{21.124}$$

To discuss asymmetries at the Z^0 resonance, we set $q^2 = m_Z^2$. The term (21.124) adds to the piece of (21.123) proportional to Q; thus it shifts the bare value of $\sin^2\theta_w$. When we include this correction, the Z^0 coupling takes the form

$$\sqrt{g^2 + g'^2}\big(T^3 - s_*^2 Q\big), \tag{21.125}$$

where

$$s_*^2 = \frac{g'^2}{g^2 + g'^2} - \frac{e}{\sqrt{g^2 + g'^2}}\frac{\Pi_{\gamma Z}(m_Z^2)}{m_Z^2}. \tag{21.126}$$

The asymmetries at the Z^0 resonance discussed in Section 20.2 are computed as ratios of these couplings. Thus, to include the oblique radiative correction to A_{LR}^f, for any light fermion species f, we reevaluate formula (20.96), using s_*^2 in place of the zeroth-order $\sin^2\theta_w$.

We might, in fact, say that s_*^2 gives an additional way to define $\sin^2\theta_w$ from observable quantities, to be compared to the definitions s_W^2 given in (21.110) and s_0^2 given in (21.111). Speaking strictly, the value of $\sin^2\theta_w$ determined by the asymmetries at the Z^0 depends on the quark or lepton quantum numbers through vertex corrections that are not included in the analysis above. However, these species-dependent corrections are small and can be systematically subtracted to define a universal s_*^2 that determines the weak interaction asymmetries of all fermion species.*

The three definitions of $\sin^2\theta_w$ all agree at zeroth order but receive different radiative corrections. If we include only the oblique corrections, it is easy to produce compact formulae for the three quantities. From (21.126), we have

$$s_*^2 = \frac{g'^2}{g^2+g'^2} - \sin\theta_w\cos\theta_w\frac{\Pi_{\gamma Z}}{m_Z^2}. \tag{21.127}$$

In the prefactor of the one-loop correction, we can ignore the distinction between the bare and renormalized values of $\sin^2\theta_w$. We can obtain a similar expression for s_W^2 by taking the ratio of the two formulae in (21.118):

$$s_W^2 = \frac{g'^2}{g^2+g'^2} - \frac{1}{m_Z^2}\left(\Pi_{WW}(m_W^2) - \frac{m_W^2}{m_Z^2}\Pi_{ZZ}(m_Z^2)\right). \tag{21.128}$$

Finally, we can evaluate the oblique corrections to $\sin^2\theta_0$ defined by (21.111). This is most readily done by writing $\delta\theta_0$ for the difference between the true and the bare value of θ_0, and then expanding (21.111) as follows:

$$2\cos 2\theta_0\,\delta\theta_0 = \frac{1}{2}\sin 2\theta_0\left[\frac{\delta\alpha}{\alpha} - \frac{\delta G_F}{G_F} - \frac{\delta m_Z^2}{m_Z^2}\right]. \tag{21.129}$$

The shifts of α, G_F, and m_Z^2 can be read from (21.120), (21.122), and (21.118). Then we can reconstruct

$$\sin^2\theta_0 = \frac{g'^2}{g^2+g'^2} + 2\sin\theta_0\cos\theta_0\delta\,\theta_0. \tag{21.130}$$

Assembling the pieces and evaluating the coefficients of the vacuum polarization diagrams to zeroth order, we obtain

$$\sin^2\theta_0 = \frac{g'^2}{g^2+g'^2} + \frac{\sin^2\theta_w\cos^2\theta_w}{\cos^2\theta_w - \sin^2\theta_w}\left[\Pi'_{\gamma\gamma}(0) + \frac{1}{m_W^2}\Pi_{WW}(0) - \frac{1}{m_Z^2}\Pi_{ZZ}(m_Z^2)\right]. \tag{21.131}$$

It is not difficult to discover that each of the equations (21.127), (21.128), and (21.131) contains ultraviolet divergences. However, if the weak interaction gauge theory is renormalizable, these divergences should cancel when we

*This is explained clearly in D. Kennedy and B. W. Lynn, *Nucl. Phys.* **B322**, 1 (1989).

compute the corrections to any zeroth-order natural relation. In the situation that we consider, renormalizability implies that the various definitions of $\sin^2\theta_w$ should differ only by expressions that are ultraviolet-finite.

We are now almost prepared to check this prediction explicitly. We can clarify the structure of the ultraviolet divergences in our relations for the various quantities $\sin^2\theta_w$ by recasting the vacuum polarization amplitudes to make more explicit the quantum numbers to which the gauge bosons couple. Recall from Eq. (20.71) that the Z boson couples to the combination of $SU(2)$ and electromagnetic quantum numbers $(T^3 - \sin^2\theta_w Q)$. Similarly, the W bosons couple to $T^{\pm}$, or, equivalently, to T^1, T^2. It is useful to break up the vacuum polarization amplitudes into terms that depend on these specific quantum numbers. We will also extract the coupling constants indicated in (20.71). Thus we replace

$$\begin{aligned}
\Pi_{\gamma\gamma} &= e^2\Pi_{QQ},\\
\Pi_{\gamma Z} &= \Big(\frac{e^2}{\sin\theta_w\cos\theta_w}\Big)\big[\Pi_{3Q} - \sin^2\theta_w\Pi_{QQ}\big],\\
\Pi_{ZZ} &= \Big(\frac{e}{\sin\theta_w\cos\theta_w}\Big)^2\big[\Pi_{33} - 2\sin^2\theta_w\Pi_{3Q} + \sin^4\theta_w\Pi_{QQ}\big],\\
\Pi_{WW} &= \Big(\frac{e}{\sin\theta_w}\Big)^2\Pi_{11},
\end{aligned} \tag{21.132}$$

where Q denotes the electric charge and $1, 2, 3$ denote the components of weak-interaction $SU(2)$.

A vacuum polarization amplitude can always be viewed as an expectation value of a pair of currents. From this viewpoint, the quantities on the right-hand side of (21.132) are expectation values of currents with definite quantum numbers. For example, Π_{33} is an expectation value of a pair of $SU(2)$ currents $J^{\mu 3}$. Acting on the standard fermions, J^{μ}_a is a left-handed current and J^{μ}_Q is a vector current.

The ultraviolet divergences in the expectation values of currents in (21.132) have the form

$$\begin{aligned}
\Pi_{33} &\sim \big(A + Bq^2\big)\log\Lambda^2,\\
\Pi_{11} &\sim \big(A + Bq^2\big)\log\Lambda^2,\\
\Pi_{3Q} &\sim \big(Bq^2\big)\log\Lambda^2,\\
\Pi_{QQ} &\sim \big(Cq^2\big)\log\Lambda^2.
\end{aligned} \tag{21.133}$$

We will demonstrate this explicitly later in this section. However, we can understand this structure from the following rough argument: Since the symmetry of the theory should be recovered at large momentum, the amplitudes Π_{33} and Π_{11}, which differ only by their orientation in the symmetry space, should have the same ultraviolet divergences. The divergence in the slope of Π_{3Q} should be related to that in the slope of Π_{33} because $Q = T^3 + Y$ and Π_{3Y} is unimportant asymptotically since $\mathrm{tr}[T^3 Y] = 0$. We pointed out

in Eq. (21.116) that Π_{3Q} and Π_{QQ} vanish at $q^2 = 0$; thus they have no q^2-independent divergences.

Now we will rewrite the two zeroth-order natural relations in (21.113) in such a way that we can apply (21.133). To do this, we take the differences of Eqs. (21.127), (21.128), and (21.131) to obtain

$$\begin{aligned} s_*^2 - \sin^2\theta_0 &= \frac{\sin^2\theta_w \cos^2\theta_w}{\cos^2\theta_w - \sin^2\theta_w}\left\{\frac{\Pi_{ZZ}(m_Z^2)}{m_Z^2} - \frac{\Pi_{WW}(0)}{m_W^2} - \Pi'_{\gamma\gamma}(0)\right. \\ &\qquad \left. - \frac{\cos^2\theta_w - \sin^2\theta_w}{\sin\theta_w\cos\theta_w}\frac{\Pi_{\gamma Z}(m_Z^2)}{m_Z^2}\right\}, \\ s_W^2 - s_*^2 &= -\frac{\Pi_{WW}(m_W^2)}{m_Z^2} + \frac{m_W^2}{m_Z^2}\frac{\Pi_{ZZ}(m_Z^2)}{m_Z^2} + \sin\theta_w\cos\theta_w\frac{\Pi_{\gamma Z}(m_Z^2)}{m_Z^2}. \end{aligned} \tag{21.134}$$

Inserting (21.132), and also using the relation $m_W = m_Z\cos\theta_w$ in the coefficients of terms already of one-loop order, we find after some algebra

$$\begin{aligned} s_*^2 - \sin^2\theta_0 &= \frac{e^2}{(\cos^2\theta_w - \sin^2\theta_w)m_Z^2}\Big\{\big[\Pi_{33}(m_Z^2) - \Pi_{11}(0) - \Pi_{3Q}(m_Z^2)\big] \\ &\qquad + \sin^2\theta_w\cos^2\theta_w\big[\Pi_{QQ}(m_Z^2) - m_Z^2\Pi'_{QQ}(0)\big]\Big\}, \\ s_W^2 - s_*^2 &= \frac{e^2}{\sin^2\theta_w m_Z^2}\big[\Pi_{33}(m_Z^2) - \Pi_{11}(m_W^2) - \sin^2\theta_w\Pi_{3Q}(m_Z^2)\big]. \end{aligned} \tag{21.135}$$

If indeed the ultraviolet divergences of the vacuum polarization integrals have the structure of (21.133), then the divergent part of each expression in brackets in (21.135) vanishes, and the weak interaction gives definite, finite predictions for the differences of s_*^2, s_W^2, and $\sin^2\theta_0$.

Computation of Vacuum Polarization Amplitudes

We can verify the divergence structure (21.133) by computing the vacuum polarization diagrams for t and b quarks explicitly. Rather than computing these one by one, it is easiest to compute, once and for all, the most general fermionic vacuum polarization amplitudes, and then to recover the amplitudes required in the previous paragraph as special cases of these.

Consider, then, the two vacuum polarization amplitudes shown in Fig. 12.13. The diagrams are built from two fermion propagators with different masses m_1 and m_2, linked by left- or right-handed currents. We call the vacuum polarization amplitude with two left-handed currents $\Pi_{LL}^{\mu\nu}(q^2)$, and that with one left and one right-handed current $\Pi_{LR}^{\mu\nu}(q^2)$. Since the vacuum polarizations depend on only one momentum and two vector indices, there is no way that they can contain an invariant involving $\epsilon^{\mu\nu\rho\sigma}$. Thus, the amplitudes with other combinations of currents are related to these by

$$\Pi_{RR}^{\mu\nu}(q^2) = \Pi_{LL}^{\mu\nu}(q^2), \qquad \Pi_{RL}^{\mu\nu}(q^2) = \Pi_{LR}^{\mu\nu}(q^2). \tag{21.136}$$

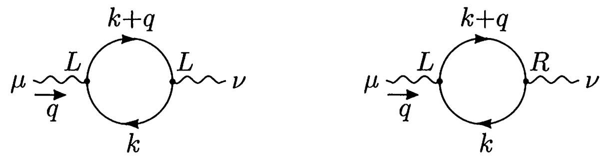
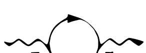


Figure 21.13. Elementary vacuum polarization amplitudes of fermionic currents.

In addition, there is no difficulty in regularizing these diagrams using dimensional regularization with an anticommuting γ^5, the regularization prescription we endorsed at the end of Section 19.4. The vacuum polarization of a vector current is reconstructed as

$$\Pi^{\mu\nu}_{VL}(q^2) = \Pi^{\mu\nu}_{LL}(q^2) + \Pi^{\mu\nu}_{RL}(q^2). \tag{21.137}$$

The vacuum polarization of purely left-handed currents is given by



$$\begin{aligned} &= (-1)\int\frac{d^4k}{(2\pi)^4}\,\mathrm{tr}\Big[(i\gamma^\mu)\Big(\frac{1-\gamma^5}{2}\Big)\frac{i(\not{k}+m_1)}{k^2-m_1^2} \\ &\qquad\cdot(i\gamma^\nu)\Big(\frac{1-\gamma^5}{2}\Big)\frac{i(\not{k}+\not{q}+m_2)}{(k+q)^2-m_2^2}\Big] \\ &= -\int\frac{d^4k}{(2\pi)^4}\,\mathrm{tr}\big[\gamma^\mu\not{k}\gamma^\nu(\not{k}+\not{q})\Big(\frac{1+\gamma^5}{2}\Big)\big]\cdot\frac{1}{(k^2-m_1^2)((k+q)^2-m_2^2)}. \end{aligned} \tag{21.138}$$

The prefactor (-1) comes from the fermion loop. There is no possible tensor structure antisymmetric in μ and ν, so we can now drop the γ^5 term. From here, the calculation proceeds as in Section 7.5. We combine denominators using

$$\frac{1}{(k^2-m_1^2)((k+q)^2-m_2^2)} = \int_0^1 dx\,\frac{1}{(\ell^2-\Delta)^2}, \tag{21.139}$$

where

$$\ell = k + xq, \qquad \Delta = xm_2^2 + (1-x)m_1^2 - x(1-x)q^2. \tag{21.140}$$

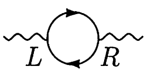
Then, integrating with dimensional regularization and following the steps leading to Eq. (7.90), we find



$$\begin{aligned} &= -\frac{4i}{(4\pi)^{d/2}}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}\big[g^{\mu\nu}\big(x(1-x)q^2 \\ &\qquad - \tfrac{1}{2}(xm_2^2+(1-x)m_1^2)\big) - x(1-x)q^\mu q^\nu\big]. \end{aligned} \tag{21.141}$$

Notice that both $\Pi^{\mu\nu}_{LL}$ and its first derivative with respect to q^2 are logarithmically divergent.

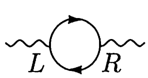
The vacuum polarization amplitude $\Pi^{\mu\nu}_{LR}$ can be obtained in a very similar fashion. From the Feynman rules,



$$= (-1)\int\frac{d^4k}{(2\pi)^4}\,\mathrm{tr}\left[(i\gamma^\mu)\left(\frac{1-\gamma^5}{2}\right)\frac{i(\not k+m_1)}{k^2-m_1^2}\cdot(i\gamma^\nu)\left(\frac{1+\gamma^5}{2}\right)\frac{i(\not k+\not q+m_2)}{(k+q)^2-m_2^2}\right]$$

$$= -\int\frac{d^4k}{(2\pi)^4}\,\mathrm{tr}\left[\gamma^\mu m_1\gamma^\nu m_2\left(\frac{1+\gamma^5}{2}\right)\right]\frac{1}{(k^2-m_1^2)((k+q)^2-m_2^2)}. \tag{21.142}$$

From here, the same manipulations as in the previous paragraph lead to



$$= -\frac{2i}{(4\pi)^{d/2}}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}\left[g^{\mu\nu}m_1m_2\right]. \tag{21.143}$$

As a check, we can use (21.141), (21.143), and (21.136), setting $m_1 = m_2 = m$, to assemble the QED vacuum polarization of vector currents. We find

$$\begin{aligned}\Pi^{\mu\nu}_{VV}(q^2) &= e^2\left[\Pi^{\mu\nu}_{LL}+\Pi^{\mu\nu}_{LR}+\Pi^{\mu\nu}_{RL}+\Pi^{\mu\nu}_{RR}\right]\\ &= \frac{-8e^2}{(4\pi)^{d/2}}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}\left[x(1-x)q^2g^{\mu\nu}-x(1-x)q^\mu q^\nu\right],\end{aligned} \tag{21.144}$$

where now $\Delta = m^2 - x(1-x)q^2$. This coincides precisely with our result from Section 7.5.

As we argued below (21.115), only the terms in the vacuum polarization amplitudes proportional to $g^{\mu\nu}$ will enter our expressions for weak-interaction radiative corrections. Thus, we can summarize the calculation of the basic vacuum polarization amplitudes by quoting the results for this leading form factor:

$$\Pi_{LL}(q^2) = \Pi_{RR}(q^2) = -\frac{4}{(4\pi)^{d/2}}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}\left[x(1-x)q^2 - \tfrac{1}{2}(xm_2^2+(1-x)m_1^2)\right];$$

$$\Pi_{LR}(q^2) = \Pi_{RL}(q^2) = -\frac{2}{(4\pi)^{d/2}}\int_0^1 dx\,\frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}}\left[m_1m_2\right]. \tag{21.145}$$

From these terms, we can assemble any desired vacuum polarization of t and b quarks in the weak-interaction gauge theory. To make use of these expressions more easily, we will expand the quantities (21.145) in the limit $d \to 4$. If we set $\epsilon = 4 - d$, the integrands of the expressions above simplify according to

$$\frac{1}{(4\pi)^{d/2}} \frac{\Gamma(2-\frac{d}{2})}{\Delta^{2-d/2}} \to \frac{1}{(4\pi)^2}\left[\frac{2}{\epsilon} - \gamma + \log(4\pi) - \log\Delta\right]. \tag{21.146}$$

Let

$$E = \frac{2}{\epsilon} - \gamma + \log(4\pi) - \log(M^2), \tag{21.147}$$

where M is an arbitrary subtraction scale. It is useful to define

$$\begin{aligned} b_0(12X) &\equiv b_0(m_1^2, m_2^2, q_X^2) = \int_0^1 dx\, \log\big(\Delta(m_1^2, m_2^2, q_X^2)/M^2\big), \\ b_1(12X) &\equiv b_1(m_1^2, m_2^2, q_X^2) = \int_0^1 dx\, x \log\big(\Delta(m_1^2, m_2^2, q_X^2)/M^2\big), \\ b_2(12X) &\equiv b_2(m_1^2, m_2^2, q_X^2) = \int_0^1 dx\, x(1-x) \log\big(\Delta(m_1^2, m_2^2, q_X^2)/M^2\big). \end{aligned} \tag{21.148}$$

The abbreviated notation will prove useful below. In (21.148), X labels a momentum scale; we will need $q_X = 0, m_W, m_Z$. Note that for equal masses,

$$b_1(11X) = \tfrac{1}{2} b_0(11X). \tag{21.149}$$

With this notation,

$$\begin{aligned} \Pi_{LL}(q_X^2) = -\frac{4}{(4\pi)^2}\Big[\big(\tfrac{1}{6}q_X^2 - \tfrac{1}{4}(m_1^2 + m_2^2)\big)E - q_X^2 b_2(12X) \\ + \tfrac{1}{2}\big(m_2^2 b_1(12X) + m_1^2 b_1(21X)\big)\Big] \end{aligned} \tag{21.150}$$

and

$$\Pi_{LR}(q_X^2) = -\frac{2}{(4\pi)^2}\big[m_1 m_2 E - m_1 m_2 b_0(12X)\big]. \tag{21.151}$$

We can now reconstruct all of the specific vacuum polarization amplitudes that appear in Eq. (21.135) in terms of divergences proportional to E and finite parts proportional to the b_i. The simplest is the expectation value of

electromagnetic currents, which is given in our present notation by

$$\Pi_{QQ}(q_X^2) = -3 \cdot \frac{8}{(4\pi)^2}\Big[\big(\tfrac{2}{3}\big)^2\big(\tfrac{1}{6}q_X^2 E - q_X^2 b_2(ttX)\big) + \big(\tfrac{1}{3}\big)^2\big(\tfrac{1}{6}q_X^2 E - q_X^2 b_2(bbX)\big)\Big]. \tag{21.152}$$

The prefactor 3 is the trace over colors. As we expect from QED, (21.152) contains a divergence only in a term proportional to q_X^2. The divergent parts of the other amplitudes are

$$\begin{aligned}
\Pi_{33}(q_X^2) &= -\frac{12}{(4\pi)^2}\cdot\tfrac{1}{2}\big[\tfrac{1}{6}q_X^2 - \tfrac{1}{4}(m_t^2+m_b^2)\big]E + \cdots,\\
\Pi_{11}(q_X^2) &= -\frac{12}{(4\pi)^2}\cdot\tfrac{1}{2}\big[\tfrac{1}{6}q_X^2 - \tfrac{1}{4}(m_t^2+m_b^2)\big]E + \cdots,\\
\Pi_{3Q}(q_X^2) &= -\frac{12}{(4\pi)^2}\cdot\tfrac{1}{2}\big[\tfrac{1}{6}q_X^2\big]E + \cdots.
\end{aligned} \tag{21.153}$$

These divergences indeed follow the pattern claimed in Eq. (21.133), and thus the predictions of the weak interaction gauge theory given in (21.135) are free of ultraviolet divergences.

The Effect of m_t

Using the notation we have developed, we can write the finite parts of the relations (21.135) in a compact form. The first relation becomes

$$\begin{aligned}
s_*^2 - \sin^2\theta_0 = \frac{3\alpha}{\pi(\cos^2\theta_w - \sin^2\theta_w)}\Big\{&\big(\tfrac{1}{4}-\tfrac{1}{3}\big)b_2(ttZ) + \big(\tfrac{1}{4}-\tfrac{1}{6}\big)b_2(bbZ)\\
&- \tfrac{1}{4}\Big(\frac{m_t^2}{m_Z^2}[b_1(ttZ) - b_1(bt0)] + \frac{m_b^2}{m_Z^2}[b_1(bbZ) - b_1(tb0)]\Big)\\
&+ 2\sin^2\theta_w\cos^2\theta_w\big(\tfrac{4}{9}[b_2(ttZ) - b_2(tt0) - m_Z^2 b_2'(tt0)]\\
&+ \tfrac{1}{9}[b_2(bbZ) - b_2(bb0) - m_Z^2 b_2'(bb0)]\big)\Big\}.
\end{aligned} \tag{21.154}$$

Similarly, the second relation becomes

$$\begin{aligned}
s_W^2 - s_*^2 = \frac{3\alpha}{\pi\sin^2\theta_w}\Big\{&\big(\tfrac{1}{4}-\tfrac{1}{3}\sin^2\theta_w\big)b_2(ttZ) + \big(\tfrac{1}{4}-\tfrac{1}{6}\sin^2\theta_w\big)b_2(bbZ)\\
&- \tfrac{1}{4}\cos^2\theta_w b_2(tbW)\\
&- \tfrac{1}{4}\Big(\frac{m_t^2}{m_Z^2}[b_1(ttZ) - b_1(btW)] + \frac{m_b^2}{m_Z^2}[b_1(bbZ) - b_1(tbW)]\Big)\Big\}.
\end{aligned} \tag{21.155}$$

Though it is now straightforward to work out the complete expressions for the relations (21.154) and (21.155), we will content ourselves here with identifying the most important term in the limit in which the t quark mass becomes large. Notice that, in each of these expressions, there are terms with

coefficients proportional to m_t^2/m_Z^2. These are easiest to understand within the simpler combination of vacuum polarization amplitudes

$$\begin{aligned}
\Pi_{11}(0) - \Pi_{33}(0) &= \frac{12}{(4\pi)^2}\frac{1}{4}\Big[m_t^2\big(b_1(tt0) - b_1(bt0)\big) + m_b^2\big(b_1(bb0) - b_1(tb0)\big)\Big] \\
&= \frac{3}{16\pi^2}\int_0^1 dx \Big\{xm_t^2 \log\frac{m_t^2}{M^2} + (1-x)m_b^2 \log\frac{m_b^2}{M^2} \\
&\qquad - (xm_t^2 + (1-x)m_b^2)\log\frac{xm_t^2 + (1-x)m_b^2}{M^2}\Big\} \\
&= -\frac{3}{16\pi^2}\int_0^1 dx \Big\{xm_t^2 \log\frac{xm_t^2 + (1-x)m_b^2}{m_t^2} + \mathcal{O}(m_b^2)\Big\} \\
&= \frac{3}{16\pi^2}\cdot\frac{1}{4}m_t^2 + \mathcal{O}(m_b^2)
\end{aligned} \tag{21.156}$$

for $m_t \gg m_b$. If m_t is also much greater than m_Z, one can find a contribution proportional to m_t^2/m_Z^2 in each of the relations (21.154), (21.155) by replacing the argument $q_X^2 = m_Z^2$ with $q_X^2 = 0$, using (21.156), and ignoring all other contributions. One can show, by detailed examination of (21.154) and (21.155), that this procedure gives the complete leading term in m_t. The result is

$$\begin{aligned}
s_*^2 - \sin^2\theta_0 &= -\frac{3\alpha}{16\pi(\cos^2\theta_w - \sin^2\theta_w)}\frac{m_t^2}{m_Z^2} + \cdots, \\
s_W^2 - s_*^2 &= -\frac{3\alpha}{16\pi\sin^2\theta_w}\frac{m_t^2}{m_Z^2} + \cdots,
\end{aligned} \tag{21.157}$$

where the omitted terms are of order α with no enhancement.

The enhancement factor m_t^2/m_Z^2 is exactly the one that we found in our study of top quark decay in Section 21.2. It reflects the fact that some electroweak couplings of the top quark are effectively proportional to λ_t, the top quark coupling to the Higgs sector, instead of simply to the weak interaction coupling g.

The complete numerical evaluation of the formulae for s_*^2 and s_W^2 is shown in Fig. 21.14. To compare the results of this section to experiment, we have included, in addition to the top quark effect, the m_t-independent one-loop corrections from loops containing W and Z bosons and light quarks and leptons. In the figure, the predictions are compared to the value of s_*^2 obtained from the measurement of the Z^0 polarization and forward-backward asymmetries and the value of s_W^2 obtained from measurement of the W boson mass.

According to the figure, the weak interaction gauge theory requires the top quark mass radiative correction (or a similar radiative correction from some other heavy particle) for its consistency with experiment. The top quark is predicted to have a mass approximately equal to 170 GeV. A recent analysis

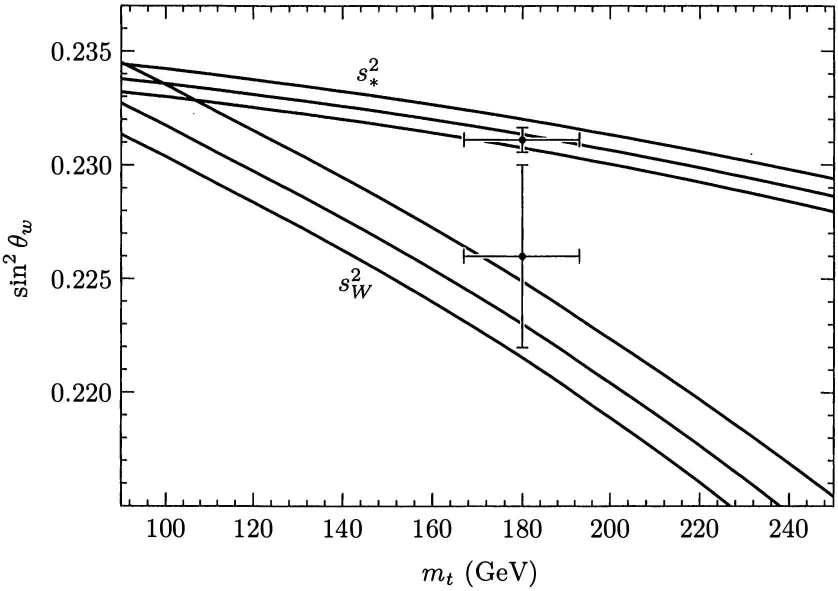


Figure 21.14. Dependence of s_*^2 and s_W^2 on the top quark mass, for fixed α, G_F, m_Z. The three curves in each group correspond to three different values of the Higgs boson mass: 100, 300, 1000 GeV from bottom to top. The curves are compared to values of s_*^2 and s_W^2, taken from the article of Langacker and Erler quoted in Table 20.1, and the CDF/D0 value of the top quark mass given in Eq. (21.159).

of all neutral current weak interaction data has given the prediction†

$$m_t = 169 \pm 24 \text{ GeV}. \tag{21.158}$$

Just as this book was being completed, the CDF and D0 experiments at Fermilab announced the observation of the production of top quark pairs in proton-antiproton scattering. From kinematic fits to events believed to contain top quarks, these experiments reported‡

$$m_t = 180 \pm 13 \text{ GeV}. \tag{21.159}$$

The discovery of the top quark in just the range required by precision electroweak measurements is quite remarkable. We can only conclude that, in the domain of weak interactions as well as those of electromagnetic, strong, and scalar interactions that we have studied earlier, the fluctuations predicted by quantum field theory make their imprint on the phenomena of Nature.

†P. Langacker and J. Erler, in *Review of Particle Properties, Phys. Rev.* **D50**, 1304 (1994).

‡F. Abe, et. al., *Phys. Rev. Lett.* **74**, 2626 (1995); S. Abachi, et. al., *Phys. Rev. Lett.* **74**, 2632 (1995).

Problems

21.1 Weak-interaction contributions to the muon $g - 2$. The GWS model of the weak interactions leads to two new contributions to the anomalous magnetic moments of the leptons. Because these contributions are proportional to $G_F m_\ell^2$, they are extremely small for the electron, but for the muon they might possibly be observable. Both contributions are larger than the contribution of the Higgs boson discussed in Problem 6.3.

(a) Consider first the contribution to the muon electromagnetic vertex function that involves a W-neutrino loop diagram. In the R_ξ gauges, this diagram is accompanied by diagrams in which W propagators are replaced by propagators for Goldstone bosons. Compute the sum of these diagrams in the Feynman-'t Hooft gauge and show that, in the limit $m_W \gg m_\mu$, they contribute the following term to the anomalous magnetic moment of the muon:

$$a_\mu(\nu) = \frac{G_F m_\mu^2}{8\pi^2\sqrt{2}} \cdot \frac{10}{3}.$$

(b) Repeat the calculation of part (a) in a general R_ξ gauge. Show explicitly that the result of part (a) is independent of ξ.

(c) A second new contribution is that from a Z-muon loop diagram and the corresponding diagram with the Z replaced by a Goldstone boson. Show that these diagrams contribute

$$a_\mu(Z) = -\frac{G_F m_\mu^2}{8\pi^2\sqrt{2}} \cdot \left(\frac{4}{3} + \frac{8}{3}\sin^2\theta_w - \frac{16}{3}\sin^4\theta_w\right).$$

21.2 Complete analysis of $e^+e^- \to W^+W^-$.

(a) Using explicit polarization vectors, work out the amplitudes for $e^+e^- \to W^+W^-$ from left- and right-handed electrons to states in which the W^+ and W^- have definite helicity. For the cases in which both W bosons have longitudinal polarization, verify that Eq. (21.99) gives the correct high-energy limit for right-handed electrons, and verify the complete expression (21.108) for left-handed electrons. For the cases in which one W is longitudinally polarized and the second is transversely polarized, show that the individual diagrams give contributions to the amplitudes that grow like $\sqrt{s}$, but that the complete amplitudes fall as $1/\sqrt{s}$.

(b) Show that the contributions to $e_L^- e_R^+ \to W^-W^+$ found in part (a) reproduce Fig. 21.10, and that the differential cross section for $e_R^- e_L^+ \to W^-W^+$ is about 30 times smaller. How many of the qualitative features of the figure can you understand physically?

21.3 Cross section for $d\bar{u} \to W^-\gamma$. Compute the amplitudes for $d\bar{u} \to W^-\gamma$ for the various possible initial and final helicities. Ignore the quark masses. In this approximation, only the annihilation amplitude from $d_L\bar{u}_R$ is nonzero. Show that the scattering amplitudes for all final helicity combinations vanish at $\cos\theta = -1/3$, where θ is the scattering angle in the center-of-mass system. Compute the differential cross section as a function of $\cos\theta$.

21.4 Dependence of radiative corrections on the Higgs boson mass.

(a) Consider the contributions to weak-interaction radiative corrections involving the physical Higgs boson h^0 of the GWS model. The couplings of the h^0 were discussed near the end of Section 20.2. Show that, if we ignore terms proportional to the masses of light fermions, the Higgs boson contributes one-loop corrections to the processes considered in Section 21.3 only through vacuum polarization diagrams. It follows that the contributions to vacuum polarization amplitudes that depend on the Higgs boson mass are gauge invariant.

(b) Draw the vacuum polarization diagrams in Feynman-'t Hooft gauge that involve the Higgs boson, and compute the dependence of the various vacuum polarization amplitudes on the Higgs boson mass m_h.

(c) Show that, for $m_h \gg m_W$, the natural relations discussed in Section 21.3 receive corrections

$$s_*^2 - s_0^2 = \frac{\alpha}{\cos^2\theta_w - \sin^2\theta_w} \frac{(1 + 9\sin^2\theta_w)}{48\pi} \log\frac{m_h^2}{m_W^2},$$

$$s_W^2 - s_*^2 = \alpha \frac{5}{24\pi} \log\frac{m_h^2}{m_W^2}.$$

The effect of varying m_h is displayed in Fig. 21.14 and is included as a theoretical uncertainty in the prediction (21.158). More accurate experiments might allow one to predict m_h from its effect on electroweak radiative corrections.

Final Project

Decays of the Higgs Boson

At the end of Section 20.2, we discussed the mystery of the origin of spontaneous symmetry breaking in the weak interactions. The simplest hypothesis is that the $SU(2) \times U(1)$ gauge symmetry of the weak interactions is broken by the expectation value of a two-component scalar field ϕ. However, since we have almost no experimental information about the mechanism of this symmetry breaking, many other possibilities can be suggested.

Eventually, this problem should be resolved by experimental observation of the particles associated with the symmetry breaking. To form incisive experimental tests, we should compute the properties expected for these particles. We saw in Section 20.2 that, if the symmetry is indeed broken by a single scalar field ϕ, the symmetry-breaking sector contributes only one new particle, a scalar h^0 called the Higgs boson. The mass m_h of this particle is unknown. However, the couplings of the h^0 to known fermions and bosons are completely determined by the masses of those particles and the weak interaction coupling constants. Thus, it is possible to compute the amplitudes for production and decay of the h^0 in some detail. More complicated models of $SU(2) \times U(1)$ symmetry breaking typically contain one or more particles that share some properties with the h^0. Thus, this study is a useful starting point for the more general analysis of experimental tests of these models.

In this Final Project you will compute the amplitudes for the Higgs boson h^0 to decay to pairs of quarks, leptons, and gauge bosons. The computations beautifully illustrate the working of perturbation theory for non-Abelian gauge fields. Those decays of the Higgs boson that involve quarks and gluons bring in aspects of QCD. Thus, this exercise reviews all of the important technical methods of Part III. Except for a question raised at the end of part (a), the problem relies only on material from unstarred sections of Part III. The material in Chapter 20 plays an essential role. Material from Chapter 21 enters the analysis only in parts (b) and (f), and the other parts of the problem (except for the final summary in part (h)) do not rely on these. If you have studied Section 19.5, you will have some additional insight into the results of parts (c) and (f), but this insight is not necessary to work the problem.

Consider, then, the minimal form of the Glashow-Weinberg-Salam electroweak theory with one Higgs scalar field ϕ. The physical Higgs boson h^0 of

this theory was discussed in Section 20.2, and we listed there the couplings of this particle to quarks, leptons, and gauge bosons. You can now use that information to compute the amplitudes for the various possible decays of the h^0 as a function of its mass m_h. You will discover that the decay pattern has a complicated dependence on the mass of the Higgs boson, with different decay modes dominating in different mass ranges. The dependences of the various decay rates on m_h illustrate many aspects of the physics of gauge theories that we have discussed in Part III.

In working this exercise, you should consider m_h as a free parameter. For the other parameters of weak-interaction theory, you might use the following values: $m_b = 5$ GeV, $m_t = 175$ GeV, $m_W = 80$ GeV, $m_Z = 91$ GeV, $\sin^2\theta_w = 0.23$, $\alpha_s(m_Z) = 0.12$.

(a) Compute first the rate for $h^0 \to f\bar{f}$, where f is a quark or lepton of the standard model. After a completely trivial computation, you should find

$$\Gamma(h^0 \to f\bar{f}) = \Big(\frac{\alpha m_h}{8\sin^2\theta_w}\Big) \cdot \frac{m_f^2}{m_W^2}\Big(1 - \frac{4m_f^2}{m_h^2}\Big)^{3/2} \cdot N_c(f),$$

where $N_c(f) = 1$ for leptons, 3 for quarks. If you have studied Chapter 18, you might improve this result for the case in which the fermion f is a quark, by computing the leading-log QCD corrections for the case $m_h \gg m_q$. Remember that the quark mass m_q is determined at the quark threshold $M^2 \sim m_q^2$.

(b) If $m_h > 2m_W$, the Higgs boson can decay to W^+W^-; if it is just a bit heavier, it can also decay to Z^0Z^0. Compute the decay widths to these final states. You can check your result in the following way: If $m_h \gg m_W$, the dominant contribution to the decay comes from production of longitudinally polarized W or Z bosons, and this contribution can be estimated at follows:

$$\Gamma(h^0 \to W^+W^-) \approx \Gamma(h^0 \to \phi^+\phi^-), \quad \Gamma(h^0 \to Z^0Z^0) \approx \Gamma(h^0 \to \phi^3\phi^3),$$

where $\phi^\pm$, ϕ^3 are the Goldstone bosons of the Higgs sector and the quantities on the right-hand sides of these relations are computed in the *ungauged* Higgs theory. Explain why this statement should be true, and verify it explicitly.

(c) The third important decay mode of the Higgs is the decay to 2 gluons. The amplitude for this decay is generated by diagrams involving quark loops. Compute these diagrams, using dimensional regularization. The diagrams will be finite, but nevertheless there is a subtle contribution which apparently depends on the regulator. Check that you have computed the amplitude correctly by verifying that it is gauge invariant. Then square the amplitude and construct the decay rate. You should find

$$\Gamma(h^0 \to 2g) = \Big(\frac{\alpha m_h}{8\sin^2\theta_w}\Big) \cdot \frac{m_h^2}{m_W^2} \cdot \frac{\alpha_s^2}{9\pi^2} \cdot \Big|\sum_q I\Big(\frac{m_h^2}{m_q^2}\Big)\Big|^2,$$

where the sum runs over all quark species and $I(m_h^2/m_q^2)$ is a form factor with the property that $I(x) \to 1$ as $x \to 0$ and $I(x) \to 0$ as $x \to \infty$. This property implies that the dominant contribution to the decay rate comes from very heavy quarks. You need not evaluate $I(x)$ explicitly at this stage; just leave it in the form of a Feynman parameter integral.

(d) The existence of the process $h^0 \to 2g$ implies the existence of the inverse process $g+g \to h^0$, which is a mechanism for production of Higgs bosons in proton-proton collisions. Using the parton model, derive a relation between the partial width $\Gamma(h^0 \to 2g)$ and the total cross section for $pp \to h^0 + X$. Compute this cross section numerically (in nanobarns) for a 30 GeV Higgs for pp collisions of center of mass energy 1–40 TeV. (1 TeV $= 10^3$ GeV.) For the purposes of this problem set (though this is not actually a good approximation) you may ignore the Q^2 dependence of the gluon distribution function and take simply

$$f_g(x) = 8 \cdot \frac{1}{x}(1-x)^7.$$

You may also set $I(m_h^2/m_t^2) = 1$; this is correct to about 10%.

(e) The final decay mode that you should consider is $h^0 \to 2\gamma$. Consider first the contribution from the loop diagrams involving quarks and leptons. Show that the result is simply expressed in terms of the form factor $I(m_h^2/m^2)$ that you derived in part (c).

(f) Next, compute the contribution to this decay amplitude from the loop diagram involving W bosons, and the various diagrams one must add to this to obtain a gauge-invariant result. It is easiest to work in Feynman-'t Hooft gauge. Add this contribution to that of very heavy quarks and leptons, each with electric charge Q_f. Your result should reduce to the following expression in the limit $m_h \ll m_W$:

$$\Gamma(h^0 \to 2\gamma) = \left(\frac{\alpha m_h}{8\sin^2\theta_w}\right) \cdot \frac{m_h^2}{m_W^2} \cdot \frac{\alpha^2}{18\pi^2} \cdot \left|\sum_f Q_f^2 N_c(f) - \frac{21}{4}\right|^2.$$

(g) Now work out the detailed behavior of the form factor $I(x)$ defined in part (c). Reduce your expression from part (c) to a one-parameter integral, then evaluate this integral numerically. Plot $I(m_h^2/m_t^2)$ over the range 50 GeV $< m_h <$ 500 GeV, and compute the decay width $\Gamma(h^0 \to 2g)$ numerically (in keV) over this range. The computation of part (f) introduces an addition form factor; compute this function in the same way.

(h) Finally, put together all the pieces. Find the branching fraction of the h^0 into each of its five major decay modes $b\bar{b}$, $t\bar{t}$, gg, W^+W^-, Z^0Z^0, for Higgs bosons of mass 50 GeV – 500 GeV.

Epilogue

Chapter 22

Quantum Field Theory at the Frontier

In this textbook we have surveyed the most important ideas of quantum field theory. Working from the basic concepts that come from fusing relativity, quantum mechanics, and fields, we have built an elaborate structure, which includes such remarkable elements as coupling constant renormalization and non-Abelian gauge fields. We have seen at many points that the strange and abstract elements of this structure actually intersect with observation and even produce explanations for unexpected aspects of the behavior of elementary particles.

In the course of our study, we have arrived at a complete theory of the strong, weak, and electromagnetic interactions of elementary particles. Each element of this theory has been described as a quantum field theory, and these quantum field theories have turned out to have very similar structure as gauge theories coupled to fermions. At various points in our discussion, we have noted that these theories have passed stringent quantitative experimental tests. We have not had space to describe the wide variety of experiments that contribute to our faith in these theories, but today almost all particle physicists consider this $SU(3) \times SU(2) \times U(1)$ gauge theory as established. In fact, most of these people refer to this theory condescendingly as 'the standard model'.

Though the best data to support the standard model have come from experiments of the past five years, the ideas behind it are much older. Most of the theoretical developments described in this book were concluded in the 1970s, a generation removed from the current frontier of physics. But this does not mean that quantum field theory is irrelevant to that frontier, any more than quantum mechanics and electrodynamics have lost their relevance after many years of exploration. On the contrary, the theory of elementary particles—like other areas of physics that depend on quantum fluctuations in continua—still holds deep mysteries, and quantum field theory remains the principal tool for exploration of these questions.

In this final chapter, we will flash forward to the present day and discuss the relevance of quantum field theory to current questions in the physics of the fundamental interactions. We will present what are, in our view, the outstanding unsolved problems of elementary particle physics and describe how quantum field theory is being used to confront these problems. Many of

these applications involve aspects of quantum field theory that are beyond the scope of this book. These include the use of quantum field theory in the regime beyond the reach of perturbation theory and the use of quantum field theory to explore the special properties of theories with higher spin and local symmetry. In these areas our discussion will be mainly qualitative, but we will give references that provide points of entry into each of these subjects.

It should be obvious that our discussion in this final chapter will express our personal opinions and by no means represents the consensus of experts in quantum field theory. In addition, any collection of 'current problems' in a rapidly developing area of research should quickly become dated. In fact, we hope the readers of this book will quickly make this chapter obsolete by solving the problems that we highlight here.

22.1 Strong Strong Interactions

One paradoxical aspect of our discussion of the strong interactions is that all of our concrete results were obtained by assuming that these interactions are weak. At large momentum transfer, we argued, this assumption is actually valid due to asymptotic freedom. Still, it is uncomfortable that we have left the most obvious questions about strongly interacting particles—for example, their masses and low-energy interactions—in a mysterious regime excluded from our analysis.

To work with QCD in the region where the strong interactions are strong, we need to answer three questions: First, how do we describe the forces that bind quarks together into hadrons? Second, what is an appropriate description of the quark-antiquark and three-quark systems bound by those forces? And finally, how do we compute scattering amplitudes and matrix elements of currents using these bound states?

At this moment, there is no derivation of the complete force between quarks from the QCD Lagrangian. Explicit calculations can be done only in the limits of weak and strong coupling. In the weak-coupling limit, the result is a Coulomb potential with an asymptotically free coupling constant. The strong coupling limit, on the other hand, gives a linear potential which confines color in the way that we described, but did not derive, at the end of Section 17.1. The derivation of this result is quite unusual and brings in a new set of mathematical methods.

So far in this book, we have not discussed a strong coupling approximation to a quantum field theory. There is a simple reason for this: In a quantum field theory in which the coupling g^2 is very large, the elementary particles or their bound states typically acquire masses that grow with g^2. For $g^2 \to \infty$, these masses become comparable to the cutoff Λ and the field theory ceases to have a local continuum description.

Wilson proposed to solve this problem in a radical way, by replacing spacetime with a lattice of discretely spaced points. Such a lattice is easiest

to visualize in Euclidean spacetime, and so we can use a functional integral over fields on a lattice to approximate Euclidean Green's functions. Such a theory can have a well-defined strong coupling limit. A theory of this type is very similar to a lattice model of a magnetic system.

In fact, we can understand this construction of a quantum field theory by using the concepts of Chapter 13. A lattice theory with fluctuating spin variables at each lattice site is described in the large by a quantum field theory of scalar fields with the symmetry of the underlying spin variables. Typically, the strong-coupling limit of the quantum field theory corresponds to the high-temperature limit of the magnet, in which the correlation length is much smaller than the lattice spacing. Decreasing the coupling constant corresponds to decreasing the temperature. Eventually, the coupling constant comes close to a fixed point of the renormalization group, and one can use this fixed point to define a limit of the lattice functional integral in which the lattice spacing is taken to zero.

To build a lattice model of the strong interactions, one needs to find a set of variables on the discrete lattice that correspond in the large to non-Abelian gauge fields. Wilson proposed that the fundamental variables for such a theory should be the line elements from one lattice vertex v_1 to a neighboring vertex v_2,

$$U(v_2, v_1) = P\exp\left[ig\int dx^\mu\, A^a_\mu t^a\right]. \tag{22.1}$$

To construct the lattice gauge theory with gauge group G, one should integrate over a finite group transformation U for each link of the lattice. Taking a product of these U matrices around a closed path, one can construct gauge-invariant observables, just as we did in Section 15.3. An appropriate Lagrangian can also be constructed as a sum of gauge-invariant products of the U matrices about elementary closed loops of the lattice.*

In a spin system, the defining property of the high-temperature phase is the exponential decay of correlations

$$\langle \vec{s}(0)\cdot\vec{s}(\mathbf{x})\rangle \sim \exp\left[-|\mathbf{x}|/\xi\right] \tag{22.2}$$

as $|\mathbf{x}|\to\infty$. The analogous property of the gauge-invariant correlation function of U matrices around a closed path P is

$$\left\langle\prod_P U\right\rangle \sim \exp\left[-A/\xi^2\right], \tag{22.3}$$

where A is the area spanned by the path. This behavior is in fact seen explicitly in the expansion of Wilson's lattice gauge theory for strong coupling. A pair of color sources that stand a distance R apart for a Euclidean time T are represented by a large rectangular loop of width R and length T. For such a

*This construction was introduced by K. Wilson, *Phys. Rev.* **D10**, 2445 (1974). The construction is described pedagogically in M. Creutz, *Quarks, Gluons, and Lattices* (Cambridge University Press, Cambridge, 1983).

loop, we can compare the result (22.3) to the expression for the energy of an excited state in Euclidean time,

$$\langle \exp[-H_E T] \rangle \sim \exp[-RT/\xi^2]. \tag{22.4}$$

Then we see that static sources of gauge charge, in the strong-coupling limit, are attracted to one another by a potential energy

$$V(R) \sim R/\xi^2 \tag{22.5}$$

at sufficiently large R. Similarly, when one introduces dynamical quarks into a lattice gauge theory and studies their properties in the strong-coupling limit, configurations with large separation of color sources are suppressed in the Euclidean functional integral by factors of the form of (22.3). The strong-coupling limit then predicts the permanent confinement of quarks into color-singlet bound states.

The argument we have just given applies equally well to gauge theories based on Abelian or non-Abelian symmetry groups. But non-Abelian gauge theories have the important additional property of asymptotic freedom. In this context, that implies that a theory with weak coupling at short distances flows to a theory with strong coupling at large distances. If we imagine integrating out short-distance degrees of freedom as we described in Section 12.1, and if there is no zero of the β function or other barrier to the renormalization group flow, we should eventually arrive at an effective theory for which the strong-coupling expansion is a good approximation. Thus, in the particular case of non-Abelian gauge theories, asymptotic freedom allows us to connect a short-distance picture based on free quarks and gluons to a large-distance picture based on color confinement.

It would be wonderful if the strong-coupling picture that we have described led to mathematical equations in continuum spacetime describing the motion of permanently confined quarks and antiquarks. Many authors have tried to write such equations by imagining the area suppression of the Wilson loop correlation function (22.3) to result from a physical surface that spans the loop. For the large rectangular loop associated with color sources, this surface can be interpreted physically as the lines of color electric flux that run from one source to the other (as in Fig. 17.1), swept out through Euclidean time. At one moment of Euclidean time, this surface can be idealized as an abstract one-dimensional excitation, often called a *string*. Unfortunately, the quantum properties of an idealized string turn out to be very complicated, since each small element of the string must be considered as an independent quantum degree of freedom. The only systems of string equations that have actually been solved have bizarre features, including unwanted massless particles. Up to now, no one has succeeded in writing an equation for the quark-confining string that is useful for quantitative calculations of quark bound states.†

†For one approach to color confinement from a picture involving Wilson loops and strings, see A. A. Migdal, *Phys. Repts.* **102**, 199 (1983).

However, the lattice regularization of a non-Abelian gauge theory suggests another approach to quantitative calculations in strong-interaction theory. By approximating QCD by a lattice gauge theory with a nonzero lattice spacing and a finite spacetime volume, we reduce the functional integral to a finite number of bounded integrations, that is, an integral over $SU(3)$ group matrices for each of the finite number of links in the lattice. A lattice of size, for example, 20^4 allows the lattice spacing to be smaller than the size of a hadron while the full size of the lattice is much larger than a hadronic radius. Then one can compute correlation functions by evaluating the integrals numerically, by the Monte Carlo method. Since the functional integral with a finite lattice spacing is related to the original functional integral with zero lattice spacing by integrating out short-distance degrees of freedom, the lattice approximation can be systematically improved by computing the short-distance effects perturbatively, using asymptotic freedom to justify a weak-coupling analysis.[‡]

This numerical method has now become the principal theoretical tool for quantitative calculations in hadron physics. This method currently gives the masses of the low-lying mesons and baryons to accuracies of 10–20%; it also allows the calculation of weak interaction matrix elements of hadrons at the 25% level. As computers become more powerful, this numerical method can be pushed to higher accuracy.

Eventually, it will be interesting to ask whether these nonperturbative numerical calculations are consistent with our precision knowledge of the perturbative region of QCD. At the time of this writing, the first such comparison has been made: We have listed in Table 17.1 a value of α_s from ψ and Υ spectroscopy. In this calculation, the experimentally determined masses of $\bar{c}c$ and $\bar{b}b$ bound states are compared to values computed numerically with lattice regularization. The comparison of these values gives the required bare coupling constant of the lattice theory, which can be converted to a value of $\alpha_s(m_Z)$ in the convention of the table. The resulting estimate for $\alpha_s(m_Z)$ does agree reasonably well with purely perturbative determinations.

What is the future of nonperturbative calculations in hadron physics? On the one hand, we expect to see further development of numerical lattice methods. These methods have hardly begun to address problems of hadron-hadron scattering and multiparticle matrix elements, and this seems an important direction for the future. In addition, these methods should eventually supply an engineering understanding of hadrons at the percent level or better. On the other hand, we hope also to see a quantitative continuum approach to hadron structure, in which dynamical quarks interact with some appropriate type of string degrees of freedom.

[‡]For an introduction to numerical lattice gauge theory, see *From Actions to Answers*, T. DeGrand and D. Toussaint, eds. (World Scientific, Singapore, 1990).

22.2 Grand Unification and its Paradoxes

If we put aside our questions about the low-energy, nonperturbative behavior of QCD, the $SU(3) \times SU(2) \times U(1)$ gauge theory gives an apparently complete description of elementary particle interactions at those energies that we have probed experimentally. But what happens beyond our current reach? Does this theory need modification, or could it continue to be valid at much higher energies?

The $SU(3) \times SU(2) \times U(1)$ gauge theory contains three independent gauge coupling constants, and the observed values of these couplings are larger for the larger components of the gauge group. This pattern can be explained by a bold hypothesis about the behavior of the gauge couplings at very high energy. If at some very large energy scale, these three couplings were equal, the values of the $SU(3)$ and $SU(2)$ couplings would increase at smaller momentum scales due to their asymptotically free renormalization group equations, while the value of the $U(1)$ coupling would decrease, resulting in the observed pattern of couplings at low energies. An even bolder hypothesis would be that the three gauge symmetries are subgroups of a single large symmetry group, which is spontaneously broken at very high energy scales. The simplest choice for this larger symmetry is $SU(5)$. In that theory, the coupling constants of $SU(3) \times SU(2) \times U(1)$ have the following relation to the underlying $SU(5)$ coupling at the scale of $SU(5)$ breaking:

$$g_5 = g_3 = g = \sqrt{\frac{5}{3}}\, g'. \tag{22.6}$$

The idea that the $SU(3) \times SU(2) \times U(1)$ gauge group is embedded in a larger simple group is known as *grand unification*; the particular choice of $SU(5)$ as the unifying group is due to Georgi and Glashow.* The observed quarks and leptons can be seen to fit neatly into an anomaly-free chiral representation of $SU(5)$; this embedding leads to a natural explanation of the fractional charges of quarks.†

Within this framework, we can extrapolate the values of the three coupling constants from the energy scale of m_Z upward. The result of this extrapolation is shown as the solid lines in Fig. 22.1. The coupling constants do come close together at very high energies, though they do not actually meet. The dashed lines in the figure show the evolution with a modified set of renormalization group equations, to be explained in Section 22.4; with this choice, the three couplings meet accurately within their current uncertainties. In any event, the evolution of coupling constants occurs on a logarithmic scale in energy, so grand unification cannot be achieved without assuming an enormous value—of order 10^{16} GeV—for the symmetry-breaking scale.

*H. Georgi and S. L. Glashow, *Phys. Rev. Lett.*, **32**, 438 (1974). The remarkable hubris of this paper makes it required reading for every student.

†For a pedagogical introduction to grand unification, see Ross (1984).

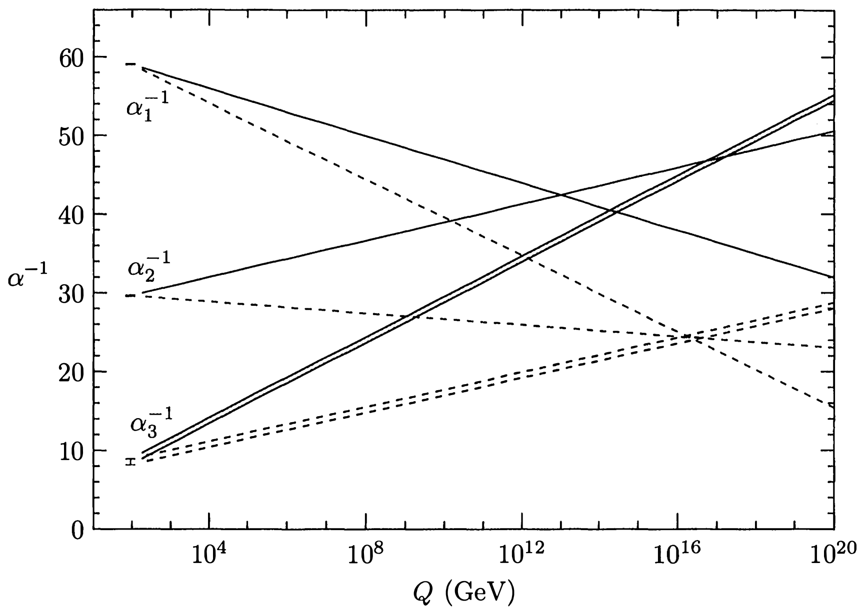


Figure 22.1. Extrapolation in energy of the coupling constants of the $SU(3) \times SU(2) \times U(1)$ gauge model, g_3, g, and $\sqrt{5/3}g'$. The solid lines are plotted using the β functions corresponding to the known set of elementary particles; the dashed lines are plotted using the β functions corresponding to a supersymmetric multiplet of particles.

The idea of a grand unification at such enormous energies raises many difficult questions, but it also suggests a wonderful opportunity. There is another enormous energy scale in quantum field theory, the scale at which the gravitational attraction of elementary particles becomes comparable to their strong, weak, and electromagnetic interactions. Conventionally, one defines the *Planck scale* as the energy for which the gravitational interaction of particles becomes of order 1:

$$m_{\text{Planck}} = (G_N/\hbar c)^{-1/2} \sim 10^{19} \text{ GeV}. \tag{22.7}$$

However, already at energies of order 10^{18} GeV, the gravitational attraction of particles is comparable to the gauge force due to the vector bosons of a grand unified theory. Though this scale is still slightly higher than the scale at which the standard model coupling constants meet, it is not unreasonable to hope that grand unification is somehow related to the unification of gravity with the forces of elementary particle physics.

On the other hand, the introduction of this large scale into physics leads to a number of conceptual problems. The first of these problems, which one meets immediately upon suggesting this extension of the standard model, is the Higgs boson mass. In our discussion at the end of Section 20.2, we came to a somewhat ambiguous conclusion about the nature of the Higgs boson. As

a part of the gauge theory of weak interactions, we need some new sector that will cause the spontaneous breaking of $SU(2) \times U(1)$. This might be supplied by the vacuum expectation value of a scalar field, or by the more complicated dynamics of a new sector of particles. At this moment, we do not know which hypothesis is to be preferred.

If $SU(2) \times U(1)$ is broken by the vacuum expectation value of an elementary scalar field, that scalar field should be part of the grand unification. This leads to a serious conceptual problem. In order to produce a vacuum expectation value of the right size to give the observed W and Z boson masses, the Higgs scalar field must obtain a negative mass term, of the size

$$-\mu^2 \sim -(100 \text{ GeV})^2. \tag{22.8}$$

Unfortunately, the $(\text{mass})^2$ of a scalar field receives additive renormalizations. In a theory with cutoff scale Λ, μ^2 can be much smaller than Λ^2 only if the bare mass of the scalar field is of order $-\Lambda^2$, and this value is canceled down to $-\mu^2$ by radiative corrections. If we envision that our theory of Nature contains the very large scales of grand unification, we must take seriously the idea that the appropriate value to take for Λ in this discussion is 10^{16} GeV or larger. This seems to require dramatic and even bizarre cancellations in the renormalized value of μ^2.

We met a situation of this type in the theory of phase transitions. At zero temperature, a ferromagnet typically has a spin expectation value of the order of the underlying atomic parameters. As the temperature is raised, or as some other variable in the system is changed, the magnetization decreases. Finally, by fine adjustment of the temperature, we can arrive at a situation where the system approaches a critical point. In the very near vicinity of this point, the expectation value of the spin field is much smaller than the value predicted from atomic parameters, and the system is described by an approximately massless continuum scalar field theory.

In statistical mechanics, this picture of the light scalar field makes sense because there is an experimenter sensitively adjusting a dial. In the theory of weak interactions, there is no one obviously making a fine adjustment that gives the $(\text{mass})^2$ of the Higgs boson a value 28 orders of magnitude or more below its natural value. Thus, it is a mystery why the Higgs boson mass should be so small compared to the grand unification scale. Particle physicists refer to this question as the *gauge hierarchy problem.*

How can one naturally arrange a Higgs field mass term to be so much smaller than the underlying mass scale of the fundamental interactions? One possible strategy would be to arrange for a symmetry of the fundamental Lagrangian that forbids the Higgs boson mass term and that is very weakly broken. This idea turns out to be very difficult to implement. To build a theory of this type, one would need to create a scalar field theory in which additive radiative corrections to the Higgs boson mass must cancel to any foreseeable

order in perturbation theory. But the Higgs mass term is very simple in form,

$$\Delta\mathcal{L} = \mu^2|\phi|^2, \tag{22.9}$$

and it is hard to imagine any principle that would keep this term from being generated by radiative corrections. There is one proposal for a symmetry with this property, but it requires the introduction of a profound principle called *supersymmetry* that entails deep modifications of fundamental physics. In particular, it requires a large number of new elementary particles, some of which should have masses below 1000 GeV, within the reach of the next generation of accelerators. We will discuss this possibility further in Section 22.4.

In this discussion, the problem of the Higgs mass stemmed from the hypothesis that the Higgs boson was an elementary particle. An alternative viewpoint, already suggested at the end of Section 20.2, is that the Higgs boson is a composite state bound by a new set of interactions. This idea also leads to observable experimental consequences, since the mass scale of these new interactions must be close to the weak interaction mass scale. In the simplest theories of this type, the symmetry breaking of the Higgs sector is modeled on the dynamical chiral symmetry breaking of the strong interactions, which we discussed in Section 19.3. The new strong interactions required by the theory lead to a spectrum of new particles with masses of about 1000 GeV.[‡] Thus, the two conflicting hypotheses on the nature of the sector that breaks $SU(2) \times U(1)$ both lead to new phenomena observable at future accelerators, and possibly even at present ones.

Just as these two different theories of the Higgs sector present completely different answers to the question of why the weak-interaction symmetry $SU(2)\times U(1)$ should be spontaneously broken, they also imply completely different answers to the question of the origin of the quark and lepton masses. In a model in which the Higgs field is elementary, the quark and lepton masses are derived from the renormalizable couplings of fermions to the Higgs field. These couplings would presumably be part of the grand unification and could be predicted only by theories that made explicit reference to the grand unification scale. In principle, the knowledge of these couplings could give us clues as to the details of the grand unification. Even if the Higgs field is composite, we cannot avoid the fact that the generation of masses for the quarks and leptons requires the breaking of $SU(2) \times U(1)$. Thus, these mass terms must arise from couplings of the quarks and leptons to the Higgs sector of interactions. In this class of models, the interactions leading to the quark and lepton masses must arise at energies close to the scale of the Higgs sector strong interaction and may eventually be observable experimentally.

From either viewpoint, it is still mysterious why the spectrum of quarks

[‡]The properties of these models of the Higgs sector, known to specialists as *technicolor* models, are described in R. Kaul, *Rev. Mod. Phys.* **55**, 449 (1983) and K. D. Lane, in *The Building Blocks of Creation*, S. Raby, ed. (World Scientific, 1993).

and leptons covers 5 orders of magnitude, from the electron at 0.5 MeV to the top quark at 175 GeV. It is also not understood what gives rise to the pattern of quark mixings encoded in the CKM matrix and the magnitude of CP violation. Even with detailed confirmation of the standard model, these questions seem today very far from solution.

The enormous mass scale of grand unification can also enter one more physical quantity, one that poses an even greater paradox than that of the Higgs boson mass. When we first quantized a field in Section 2.3, we discovered that the energy density of the vacuum in free scalar field theory received an infinite positive contribution from the zero-point energies of the various modes of oscillation. With a cutoff scale Λ, this zero-point energy is given roughly by

$$\langle 0| H |0\rangle \sim \Lambda^4. \tag{22.10}$$

At many other points in our discussion, we found similarly large contributions to the vacuum energy. The filling of the Dirac sea in the quantization of the free fermion theory led to a downward shift in the vacuum energy with a similar ultraviolet divergence. Spontaneous symmetry breaking gives a finite but still possibly large shift in the vacuum energy density,

$$\Delta \langle 0| H |0\rangle \sim -cv^4, \tag{22.11}$$

with dimensionless c, for a field vacuum expectation value v. The spontaneous breaking of the weak interaction $SU(2)\times U(1)$ symmetry and of the strong interaction chiral symmetry both would be expected to shift the vacuum energy density in this way.

In elementary particle physics experiments, this shift of the vacuum energy is unobservable. Experimentally measured particle masses, for example, are energy differences between the vacuum and certain excited states of H, and the absolute vacuum energy cancels out in the calculation of these differences. However, there is a way that the absolute vacuum energy could potentially be observed, through the coupling of the vacuum energy to gravity. According to Einstein, the energy-momentum tensor of matter $\Theta^{\mu\nu}$ is the source of the gravitational field. A vacuum energy density $\langle 0| H |0\rangle = \lambda$ contributes to this source a term

$$\Theta^{\mu\nu} = N\big(\Theta^{\mu\nu}\big) + \lambda g^{\mu\nu}, \tag{22.12}$$

where the first term on the right is subtracted to have zero vacuum expectation value. The vacuum energy term has the form of Einstein's cosmological constant and thus potentially affects the expansion of the universe.

In fact, measurements of the cosmological expansion exclude a large cosmological constant. The current limit is

$$\lambda < 10^{-29}\ \text{g/cm}^3 \sim (10^{-11}\ \text{GeV})^4. \tag{22.13}$$

We have no understanding of why λ is so much smaller than the vacuum energy shifts generated in the known phase transitions of particle physics, and so much again smaller than the underlying field zero-point energies. The

discrepancy in λ between the experimental bound (22.13) and naive intuition is 120 orders of magnitude! The solution to this problem may come from one of many sources. It may be that the formalism of the quantum field theory of gravity requires that the vacuum energy be subtracted from the energy-momentum tensor that appears in Einstein's equations of gravity. It may be that there is a new physical mechanism coming from particle physics or from gravity itself that sets the total vacuum energy to zero. Or it may be that the overall scale of energy-momentum is genuinely ambiguous and is set by a cosmological boundary condition. At this moment, all of these possibilities are just guesses. All we know for certain is that the unification of quantum field theory and gravity cannot be straightforward, that there is some important concept still missing from our current understanding.*

22.3 Exact Solutions in Quantum Field Theory

From the idea of grand unification, with its great promise and mystery, we turn to the study of model quantum field theories that are so simple that they can be solved exactly. Throughout this book, we have stressed the intrinsic complexity of quantum field theory and the importance of using perturbation theory as a replacement for exact knowledge. But there are a variety of quantum field theories for which exact solutions are known. In this section, we will describe some of these and review the insights we have gained from them.

In searching for exact solutions to quantum field theory models, there is no reason to restrict our attention to four-dimensional spacetime. In fact, we have seen examples of two-dimensional theories with similar complexity of renormalization and short-distance behavior. At the same time, these theories occupy a one-dimensional space, and their degrees of freedom can be visualized as links in a chain. This allows some powerful simplifications.

In our discussion of the axial anomaly in two dimensions in Section 19.1, we showed that the photon of two-dimensional massless QED becomes a massive boson. More detailed examination of this theory shows that this boson is a noninteracting particle. The theory is originally formulated in terms of fermions, interacting through Coulomb forces. However, it is possible to exactly rewrite the theory as a theory of a scalar field that creates and destroys fermion-antifermion pairs. Heuristically, a particle and an antiparticle moving down the light-cone in one-dimensional space do not separate and therefore comprise one bosonic degree of freedom. In a wide class of models, the bosonic theories rewritten in this way are free-field theories. A remarkable model of this type is the *Thirring model*,

$$\mathcal{L} = \bar{\psi} i \partial\!\!\!/ \psi - \frac{g}{2} \bar{\psi}\gamma^\mu\psi\bar{\psi}\gamma_\mu\psi, \tag{22.14}$$

*The cosmological constant problem and a variety of unsuccessful solutions are reviewed in S. Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989).

in two dimensions. In this model, the replacement of the fermion field by a boson field leads to a free field theory. Using this field theory, one can compute correlation functions of fermion bilinears explicitly and show directly that these operators have anomalous dimensions. In renormalization-group language, the model contains a line of fixed points parametrized by the coupling constant g.†

A more general class of two-dimensional models can be solved by visualizing them in a Hamiltonian picture as a one-dimensional chain of coupled field operators. The prototype of this method is a problem in the statistical mechanics of magnets, the one-dimensional chain of coupled spins. Consider a long chain of N discrete sites, with a spin-1/2 system at each site. The Pauli sigma matrices $\boldsymbol{\sigma}_i$ act on the two-dimensional Hilbert space at the site i. The Hamiltonian for the spin chain is then

$$H = \sum_i \big(-J\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1}\big). \tag{22.15}$$

Since

$$\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1} = 2(\sigma_i^+\sigma_{i+1}^- + \sigma_i^-\sigma_{i+1}^+) + \sigma_i^3\sigma_{i+1}^3, \tag{22.16}$$

this Hamiltonian conserves the number of up spins. The state with all spins down is an eigenstate of the Hamiltonian, and the states with one spin up in a state of definite momentum are also eigenstates. In 1934, Bethe analyzed the problem of two spins up and computed their S-matrix. He then discovered an amazing fact, that by multiplying the S-matrices for the two-spin problem, he could find the exact eigenstates of the Hamiltonian for any number of spins up. By considering $N/2$ spins up, he found the ground state of the system. This technique, now known as *Bethe's ansatz*, has been used to solve a wide variety of one-dimensional problems in condensed matter physics and quantum field theory. For example, this technique has been used by Andrei and Lowenstein to solve the Gross-Neveu model presented in Problem 11.3 and to demonstrate that the spectrum of this model has the properties expected from asymptotic freedom.‡

Even where it is not possible to solve a model for all values of its parameters, it is sometimes possible to find exact information about two-dimensional models at points where they contain massless fields. It is well known that a variety of classical two-dimensional partial differential equations can be solved by exploiting techniques of complex variables. For example, the two-dimensional Laplace equation $\nabla^2\phi = 0$ is invariant to conformal mappings $z \to w(z)$,

† For an introduction to these models, see S. Coleman, *Phys. Rev.* **D11**, 2088 (1975), *Ann. Phys.* **101**, 239 (1976).

‡ For an introduction to Bethe's ansatz and its generalizations, see N. Andrei, K. Furuya, and J. H. Lowenstein, *Rev. Mod. Phys.* **55**, 331 (1983), L. D. Faddeev, in *Recent Advances in Field Theory and Statistical Mechanics*, J. B. Zuber and R. Stora, eds. (North-Holland, Amsterdam, 1984), and R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).

where $z = x + iy$. Two-dimensional quantum field theories with massless particles often have this conformal symmetry at the classical level, though generically it is anomalous. In special systems, however, these anomalies vanish and the quantum theory is invariant to conformal mapping. These theories typically contain operators with anomalous dimensions, indicating that each such theory is a new, nontrivial fixed point of the renormalization group. The conformal symmetry of the theory can be used to compute these anomalous dimensions.

As an example of this class of theories, consider the two-dimensional nonlinear sigma model in which the basic field is not a unit vector, as we discussed in Section 13.3, but rather a unitary matrix of a Lie group G. The Lagrangian of this theory is

$$\mathcal{L} = \frac{1}{4g^2} \int d^2x \; \mathrm{tr}\big[\partial_\mu U^\dagger \partial^\mu U\big]. \tag{22.17}$$

Like the theory of Section 13.3, this model is asymptotically free. However, Witten has shown that, by adding to this Lagrangian a particular perturbation of a rather complicated form first written by Wess and Zumino, one can find a fixed point of the renormalization group with manifest $G \times G$ global symmetry. This theory is conformally invariant, and all operator correlation functions can be computed using the conformal symmetry.*

One result of the nonperturbative exploration of quantum field theory was the realization that field theories can contain particle states that are not simply related to the quanta of the original fields. In the weak-coupling limit of a quantum field theory, such new states can appear as new solutions of the classical field equations. Consider, for example, ϕ^4 theory in two dimensions in the broken-symmetry phase. The equation of motion is

$$\frac{\partial^2}{\partial t^2}\phi - \frac{\partial^2}{\partial x^2}\phi - \mu^2\phi + \lambda\phi^3 = 0. \tag{22.18}$$

Treating this equation as a classical partial differential equation, we can find the time-independent solution

$$\phi(x) = \frac{\mu}{\sqrt{\lambda}} \tanh\frac{x\mu}{\sqrt{2}}. \tag{22.19}$$

This is a field configuration that begins in one well of the potential at $x = -\infty$ and crosses over to the other well as $x \to +\infty$. This solution has an energy of order μ/λ, larger by a factor of $1/\lambda$ than the mass of a ϕ quantum. Since the original equation (22.18) was Lorentz-covariant, the boosts of this solution must also be solutions to the classical partial differential equation. It is natural to suggest that, in the ϕ^4 quantum field theory, these solutions form a new set of massive particles. Such solutions, and the particles corresponding to

*For an introduction to conformally invariant two-dimensional quantum field theories, see P. Ginsparg, in *Fields, Strings, Critical Phenomena*, E. Brezin and J. Zinn-Justin, eds. (North-Holland, Amsterdam, 1989).

them, are often called *solitons*, borrowing a more specialized term from the literature on two-dimensional partial differential equations.†

Many examples are now known of particles that are associated in this way with classical solutions of a quantum field theory. In theories with spontaneously broken symmetry, the appearance of such particles is often related to the topology of the set of vacuum states; the ϕ^4 theory above gives a simple example of this relation. These examples are not limited to two dimensions but can also occur in theories that are potentially realistic. Such solutions can have magical properties. One interesting example is found in the $SU(2)$ gauge theory with a Higgs scalar field in the vector representation, the Georgi-Glashow model considered in Section 20.1. 't Hooft and Polyakov showed that this theory has a classical solution in which the Higgs field ϕ_a has the form

$$\phi_a(\mathbf{x}) = f(|\mathbf{x}|)x_a. \tag{22.20}$$

They showed that, when the gauge theory is interpreted as a unified model of weak and electromagnetic interactions, this solution is a magnetic monopole! In addition, particles that arise as heavy classical states in the weak coupling limit can have a more intricate relation to the dynamics of the theory when the coupling is increased. For example, in theories of the type of two-dimensional QED or the Thirring model in which fermions can be replaced by bosons, a weak-coupling limit is obtained by adding to the theory a large fermion mass. Then the original fermions are recovered from the bosonic representation of the theory as classical solutions very similar to that given in (22.19).

In some theories, one can find classical solutions of the Euclidean field equations. These solutions, called *instantons*, are localized in Euclidean time as well as in space. Thus, they are interpreted as quantum processes that modify the effective Hamiltonian of a quantum field theory. The most famous example of an instanton is found in four-dimensional non-Abelian gauge theories. It was shown by 't Hooft that this field configuration leads to a quantum process that violates the conservation of the $U(1)$ axial current in QCD. We have explained in Section 19.3 that this violation of current conservation is exactly what is needed to explain the spectrum of light mesons in QCD.

There is probably much more to be learned, especially about the strong-coupling behavior of gauge theories, by deeper analysis of the classical solutions to the field equations, and of the interrelations of the many exactly or partially solvable two-dimensional field theories.

---

†For an introduction to the use of solutions of the classical field equations in the analysis of problems in field theory, see S. Coleman (1985), Chaps. 6 and 7, and R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).

22.4 Supersymmetry

Among the properties that a quantum field theory might possess to make it more beautiful or more mathematically tractable, there is one higher symmetry with particularly far-reaching implications. This is a symmetry that relates fermions and bosons, known (without hyperbole) as *supersymmetry*. In this section, we will introduce some of the purely mathematical consequences of supersymmetry, and then discuss the question of whether the true field equations of Nature could be supersymmetric.

A generator of supersymmetry is an operator that commutes with the Hamiltonian and converts bosonic into fermionic states. Such an operator must carry half-integer spin, in the simplest case spin 1/2. Let Q_α, with $\alpha = 1, 2$, be the left-handed spinor components of this operator. Their Hermitian conjugates, $Q^\dagger_\beta$, form a right-handed spinor. The anticommutator $\{Q_\alpha, Q^\dagger_\beta\}$ is a 2×2 matrix with positive diagonal elements; thus it cannot vanish. This matrix commutes with H but transforms nontrivially under Lorentz transformations. A Lorentz-covariant expression for this anticommutator is

$$\{Q_\alpha, Q^\dagger_\beta\} = 2\sigma^\mu_{\alpha\beta} P^\mu, \tag{22.21}$$

where P^μ is a conserved vector quantity. Such quantities are severely restricted; a theorem of Coleman and Mandula states that, if a quantum field theory in more than two dimensions has a second conserved vector quantity in addition to the energy-momentum 4-vector, the S-matrix equals 1 and no scattering is allowed. Thus the only possible choice for P^μ in Eq. (22.21) is the total energy-momentum. The Coleman-Mandula theorem also rules out any higher-spin conservation laws. This eliminates the possibility that a supersymmetry generator could have spin 3/2 or higher. The most general possibility is a collection of spin-1/2 operators with the anticommutation relations

$$\{Q^i_\alpha, Q^{j\dagger}_\beta\} = 2\delta^{ij}\sigma^\mu_{\alpha\beta} P^\mu, \tag{22.22}$$

with $i, j = 1, \ldots, N$. In the following discussion, we will mainly consider only the simplest case, $N = 1$.[‡]

The algebra (22.22) of conserved quantities has profound conseqences for the theory. Since the right-hand side of (22.22) is the total energy-momentum, it involves every field in the theory. To reproduce this algebra, the left-hand side must also involve every field. The representations of this algebra pair every bosonic state with a fermionic state at the same energy, and vice versa. If supersymmetry is an exact symmetry of the quantum field theory, it must act on every sector of the theory. In a realistic model, even the gravitational field must have a fermionic partner. This means that Einstein's equations of gravity must be generalized to a new set of geometrical equations that involve a fermionic (spin-3/2) field.

[‡]An excellent introduction to the formalism of supersymmetry is J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton University Press, 1983).

The first consequences of making a quantum field theory supersymmetric are easy to understand. For every (complex) scalar field, one must introduce a chiral fermion field. The self-interactions of the bosonic fields are related to the interactions of these fields with the fermions; for example, a possible interaction Lagrangian with coupling constant λ is

$$\Delta\mathcal{L} = -\lambda^2|\phi|^2 - \tfrac{1}{2}\lambda\psi^T\sigma^2\psi. \tag{22.23}$$

We have written a more general supersymmetric Lagrangian in Problem 3.5. Similarly, for every gauge field, one must introduce a chiral fermion in the adjoint representation of the gauge group. This fermion, called the *gaugino*, mediates interactions of the scalar fields with their fermionic partners whose strength is given by the gauge coupling g.

The special relation between the bosonic and fermionic interactions leads to great simplifications in the renormalization of supersymmetric theories. Some of these simplifications can be anticipated. Since supersymmetry requires that each scalar particle have a fermionic partner of the same mass, these particles must have the same mass renormalization. But we have seen that the fermion mass is multiplicatively renormalized and thus is only logarithmically divergent, while a scalar mass term is additively renormalized and thus can be quadratically divergent. Supersymmetry must imply that the quadratic divergences of scalar mass terms automatically vanish. In fact, these cancellations occur in every order of perturbation theory, with loop diagrams involving bosons canceling against diagrams with virtual fermions. To see another simplification required by supersymmetry, take the vacuum expectation value of the anticommutation relation (22.21). The vacuum state has zero momentum: $P^i\left|0\right\rangle = 0$. If the vacuum state is supersymmetric, $Q_\alpha\left|0\right\rangle = Q_\beta^\dagger\left|0\right\rangle = 0$. Then Eq. (22.21) implies

$$\left\langle 0\right| H \left|0\right\rangle = 0. \tag{22.24}$$

We have noted already that bosonic fields give positive contributions to the vacuum energy through their zero-point energy, and fermionic fields give negative contributions. We now see that, in a supersymmetric model, these contributions cancel exactly, not only at the leading order but to all orders in perturbation theory.

Deeper examination of supersymmetric theories leads to additional, and quite unexpected, cancellations in renormalization theory. For example, one can show that the coupling constants in scalar-fermion self-interactions, such as λ in (22.23), are renormalized only through field strength renormalizations. Thus the relative size of two different scalar interactions remains unchanged. If a particular type of renormalizable interaction is omitted, it cannot be generated by renormalization, in contrast to the case in ordinary field theory. The simplest supersymmetry does not constrain the renormalization of gauge couplings, but higher supersymmetries can have a profound effect: In $N = 2$ supersymmetric models, the β function vanishes if the leading-order term is

arranged to be zero. In $N = 4$ supersymmetric models, this cancellation is automatic and $\beta(g) = 0$ exactly. These models give examples of four-dimensional quantum field theories with no ultraviolet divergences.*

Supersymmetry thus endows a quantum field theory with remarkable, even magical properties. But is it possible that the true equations of Nature could possess such a high degree of symmetry? Since we are certain that there is no charged boson with the same mass as the electron, we know that supersymmetry cannot be an exact symmetry of Nature. But it is tempting to guess that it might be a spontaneously broken symmetry of the underlying equations.

In fact, this conjecture has fruitful consequences for the grand unified theories that we discussed in Section 22.2. The problem of the Higgs boson mass that we highlighted in that section has an elegant solution in supersymmetry models. In a supersymmetric version of the standard model, the Higgs field is one of a large number of scalar fields with various $SU(3) \times SU(2) \times U(1)$ quantum numbers. For all of these scalar fields, the mass terms get only a logarithmic multiplicative renormalization. If supersymmetry were broken in such a way as to give mass differences of a few hundred GeV between the observed fermionic quarks and leptons and their scalar partners, one would also find a Higgs boson $(\text{mass})^2$ of the correct size. There are good reasons, which follow from more detailed properties of the theory, why it is the Higgs field, rather than some other scalar field, that obtains a vacuum expectation value.†

If this set of ideas is correct, the scalar partners of quarks and leptons would be light enough to be discovered experimentally in the near future. In that case, these scalar particles and the fermionic partners of gauge bosons would affect the renormalization of coupling constants between present energies and the grand unification scale. This might potentially disturb the prospects for grand unification, but, instead, it improves them: the dashed lines of Fig. 22.1, with a more impressive meeting of the three coupling constants, were generated by replacing the conventional β functions with ones including the supersymmetric partners.

The last of the problems discussed in Section 22.2 is also ameliorated by the introduction of supersymmetry. In a grand unified theory with broken supersymmetry, those momentum scales that are much larger than the mass differences of supersymmetry partners give no contribution to the vacuum energy. Thus the natural size of the cosmological constant in these theories is $\lambda \sim (100 \text{ GeV})^4$. This reduces the cosmological constant problem to a discrepancy of 50 orders of magnitude—but this is not nearly enough.

---

*Supersymmetric models with vanishing β function are reviewed by P. West, in *Shelter Island II*, R. Jackiw, N. N. Khuri, S. Weinberg, and E. Witten, eds. (MIT Press, Cambridge, 1985).

†Supersymmetric models of quarks and leptons, and their observable consequences, are reviewed in H. P. Nilles, *Phys. Repts.* **110**, 1 (1984), and in H. E. Haber and G. L. Kane, *Phys. Repts.* **117**, 75 (1985).

It is an exciting prospect that supersymmetric partners of the particles of the standard model might soon be seen in experiments. What we anticipate, in any event, is that the experiments of the next generation will make a definite choice between this hypothesis for the nature of the Higgs sector and the other possibilities discussed in Section 22.2. Either way, we will have advanced our knowledge one step toward the truly fundamental equations.

22.5 Toward an Ultimate Theory of Nature

What are these fundamental equations? Do they involve quantum field theory, or some very different organizing principle? Any answer to this question must be completely speculative. Nevertheless, there are some principles, and an example, that can guide this search.

For all the attention we have given in this book to the basic interactions of particle physics, we have given very little attention to gravity. In part, this is because the quantum theory of gravity has no known observational consequences. But it is also true that the quantum theory of gravity is still ill-formed and uncertain. If gravity is treated as a weak-coupling field theory with Feynman diagrams, one quickly finds that the divergences of these diagrams render the theory nonrenormalizable. This is no surprise, because gravity is a theory in which the coupling constant has inverse mass dimensions, with the mass scale m_{Planck} given by (22.7). In our general philosophy of renormalization, all of the complexity of this theory should be concentrated at the scale m_{Planck}.

At the scale where quantum fluctuations of the gravitational field are important, we must expect profound changes in physics. If these changes occur within the context of quantum field theory, they will at the least entail fluctuating spacetime geometry and topology. But it seems equally probable that quantum field theory will actually break down at this scale, with continuous spacetime replaced by a new discrete or nonlocal geometry.

One particular model for the behavior of spacetime at very small distances is *string theory*, the dynamics of abstract one-dimensional extended objects. In Section 22.1, we mentioned that such objects seemed to occur naturally in attempts to describe quark confinement in QCD, but that the detailed properties of these objects made them unsuitable for strong interaction phenomenology. Among the disappointing properties of these systems were the appearance of massless spin-2 states of the string, and a constraint that the dimension of spacetime must be increased unless the spectrum of the theory contained many massless spin-1 states. In 1974, Scherk and Schwarz made the remarkable suggestion that string theory was a correct mathematical description of a different problem, the unification of elementary particle interactions with gravity. They interpreted the spin-2 quantum as the graviton and the spin-1 quanta as gauge bosons of a gauge theory.[‡] A decade later,

[‡] J. Scherk and J. H. Schwarz, *Nucl. Phys.* **B81**, 118 (1974).

Green and Schwarz put this conjecture on a firmer footing by showing that a particular string theory could be interpreted as a grand unified theory in ten spacetime dimensions, with all gravitational and gauge Ward identities automatically satisfied and all anomalies automatically canceling. Since that time, many other solutions to the constraint equations of string theory have been found, some of which correspond to unified models of gauge interactions and gravity in four dimensions. These models can naturally incorporate supersymmetry and, under that condition, give ultraviolet-finite results for all scattering amplitudes, including those of gravitons.*

String theories solve the ultraviolet divergence problems of quantum field theory by rejecting the idea that elementary particles are pointlike objects with contact interactions. Rather, in string theory, quarks, leptons, gauge bosons, and gravitons are extended loops of string excitation which thus interact nonlocally. Since particles cannot be definitely localized, spacetime itself takes on a nonlocal character. In some sense, distances much less than the Planck length $1/m_{\text{Planck}}$ do not exist in the string description of gravity. As yet, it is not clear how to understand intuitively the sort of geometry that string theory requires. This mathematical problem is now being actively investigated.

If indeed the truly fundamental geometry of Nature is nonlocal, discrete, or discontinuous in some other way, then the grand program for the fundamental interactions that we have set forth in this book must be altered in an essential way. The most elementary equations of Nature will not be gauge-invariant quantum field theories, but instead theories built from very different elements. Even the principles of model construction will be different from those based on gauge and Lorentz invariance that we have discussed here.

On the other hand, quantum field theory will still play an essential role in the interpretation of this structure. All of the processes we now observe, even the elementary particle processes at the highest energies currently accessible, occur over distances 15 orders of magnitude larger than the sizes of the strings or other fluctuating structures that appear in the underlying equations. The relation of experimental observations to these fundamental structures is thus very similar to the relation of macroscopic observations to the underlying atomic structure of matter. In the study of matter, we use a classical, Newtonian description of atoms to bridge this gap and to relate the properties of gases, liquids, and solids to underlying atomic properties. We might say that the quantum theory of atoms gives rise to a set of effective Newtonian equations that is extremely powerful in the macroscopic domain. Especially in the theory of gases, this Newtonian description was also used as a tool to realize the existence of atoms and to derive their properties.

---

*A technical introduction to string theory and its use in building unified models has been given by M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory*, 2 vols. (Cambridge University Press, 1987).

Similarly, whatever the nature of Planck-scale physics, it leads to some effective continuum quantum field theory. This quantum field theory might well be an accurate approximation to the underlying physics already at distances of 100 Planck lengths, corresponding to momenta of 10^{17} GeV. From here to the scale of weak interactions, and from there up to the wavelength of light, and from there to the size of the universe, quantum field theory can be treated as the basic framework for the equations of physics. By recognizing the symmetries of the particular set of field equations that Nature has provided us, we can learn to compute all of the details of physical processes over this whole enormous domain. And, by contemplating the origin of these symmetries, perhaps we will also be able to see through to the next level and unlock the true structure of spacetime.

Appendix

Reference Formulae

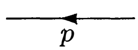
This Appendix collects together some of the formulae that are most commonly used in Feynman diagram calculations.

A.1 Feynman Rules

In all theories it is understood that momentum is conserved at each vertex, and that undetermined loop momenta are integrated over: $\int d^4p/(2\pi)^4$. Fermion (including ghost) loops receive an additional factor of (-1), as explained on page 120. Finally, each diagram can potentially have a symmetry factor, as explained on page 93.

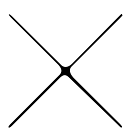
ϕ^4 theory: $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4$

Scalar propagator:



$$= \frac{i}{p^2 - m^2 + i\epsilon} \tag{A.1}$$

ϕ^4 vertex:



$$= -i\lambda \tag{A.2}$$

External scalar:

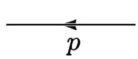


$$= 1 \tag{A.3}$$

(Counterterm vertices for loop calculations are given on page 325.)

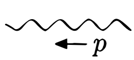
Quantum Electrodynamics: $\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi - \frac{1}{4}(F_{\mu\nu})^2 - e\bar{\psi}\gamma^\mu\psi A_\mu$

Dirac propagator:



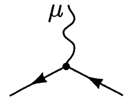
$$= \frac{i(p\!\!\!/ + m)}{p^2 - m^2 + i\epsilon} \tag{A.4}$$

Photon propagator:



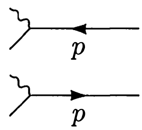
$$= \frac{-ig_{\mu\nu}}{p^2 + i\epsilon} \tag{A.5}$$

(Feynman gauge; see page 297 for generalized Lorentz gauge.)

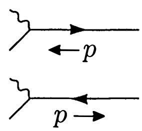
QED vertex: 

$= iQe\gamma^\mu$ (A.6)

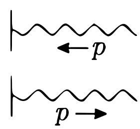
($Q = -1$ for an electron)

External fermions: 

$= u^s(p)$ (initial)
$= \bar{u}^s(p)$ (final) (A.7)

External antifermions: 

$= \bar{v}^s(p)$ (initial)
$= v^s(p)$ (final) (A.8)

External photons: 

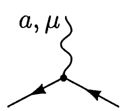
$= \epsilon_\mu(p)$ (initial)
$= \epsilon^*_\mu(p)$ (final) (A.9)

(Counterterm vertices for loop calculations are given on page 332.)

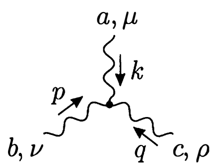
Non-Abelian Gauge Theory:

$$\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi - \tfrac{1}{4}(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)^2 + gA^a_\mu \bar{\psi}\gamma^\mu t^a \psi$$
$$- gf^{abc}(\partial_\mu A^a_\nu)A^{\mu b}A^{\nu c} - \tfrac{1}{4}g^2(f^{eab}A^a_\mu A^b_\nu)(f^{ecd}A^{\mu c}A^{\nu d})$$

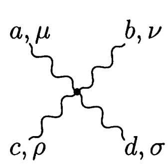
The fermion and gauge boson propagators are the same as in QED, times an identity matrix in the gauge group space. Similarly, the polarization of external particles is treated the same as in QED, but each external particle also has an orientation in the group space.

Fermion vertex: 

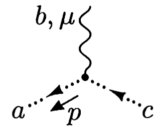
$= ig\gamma^\mu t^a$ (A.10)

3-boson vertex: 

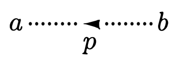
$$= gf^{abc}\big[g^{\mu\nu}(k-p)^\rho + g^{\nu\rho}(p-q)^\mu + g^{\rho\mu}(q-k)^\nu\big] \quad \text{(A.11)}$$

4-boson vertex: 

$$= -ig^2\big[f^{abe}f^{cde}(g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ace}f^{bde}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\sigma}g^{\nu\rho}) + f^{ade}f^{bce}(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma})\big] \quad \text{(A.12)}$$

Ghost vertex: 

$$= -gf^{abc}p^{\mu} \tag{A.13}$$

Ghost propagator: 

$$= \frac{i\delta^{ab}}{p^2 + i\epsilon} \tag{A.14}$$

(Counterterm vertices for loop calculations are given on pages 528 and 532.)

Other theories. Feynman rules for other theories can be found on the following pages:

Yukawa theory	page 118
Scalar QED	page 312
Linear sigma model	page 353
Electroweak theory	pages 716, 743, 753

A.2 Polarizations of External Particles

The spinors $u^s(p)$ and $v^s(p)$ obey the Dirac equation in the form

$$\begin{aligned} 0 &= (\not{p} - m)u^s(p) = \bar{u}^s(p)(\not{p} - m) \\ &= (\not{p} + m)v^s(p) = \bar{v}^s(p)(\not{p} + m), \end{aligned} \tag{A.15}$$

where $\not{p} = \gamma^\mu p_\mu$. The Dirac matrices obey the anticommutation relations

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}. \tag{A.16}$$

We use a chiral basis,

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \qquad \gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{A.17}$$

where

$$\sigma^\mu = (1, \boldsymbol{\sigma}), \qquad \bar{\sigma}^\mu = (1, -\boldsymbol{\sigma}). \tag{A.18}$$

In this basis the normalized Dirac spinors can be written

$$u^s(p) = \begin{pmatrix} \sqrt{p\cdot\sigma}\,\xi^s \\ \sqrt{p\cdot\bar{\sigma}}\,\xi^s \end{pmatrix}, \qquad v^s(p) = \begin{pmatrix} \sqrt{p\cdot\sigma}\,\eta^s \\ -\sqrt{p\cdot\bar{\sigma}}\,\eta^s \end{pmatrix}, \tag{A.19}$$

where ξ and η are two-component spinors normalized to unity. In the high-energy limit these expressions simplify to

$$u(p) \approx \sqrt{2E}\begin{pmatrix} \frac{1}{2}(1 - \hat{p}\cdot\boldsymbol{\sigma})\xi^s \\ \frac{1}{2}(1 + \hat{p}\cdot\boldsymbol{\sigma})\xi^s \end{pmatrix}, \qquad v(p) \approx \sqrt{2E}\begin{pmatrix} \frac{1}{2}(1 - \hat{p}\cdot\boldsymbol{\sigma})\eta^s \\ -\frac{1}{2}(1 + \hat{p}\cdot\boldsymbol{\sigma})\eta^s \end{pmatrix}. \tag{A.20}$$

Using the standard basis for the Pauli matrices,

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{A.21}$$

we have, for example, $\xi^s = \binom{1}{0}$ for spin up in the z direction, and $\xi^s = \binom{0}{1}$ for spin down in the z direction. For antifermions the physical spin is opposite to that of the spinor: $\eta^s = \binom{1}{0}$ corresponds to spin *down* in the z direction, and so on.

In computing unpolarized cross sections one encounters the polarization sums

$$\sum_s u^s(p)\bar{u}^s(p) = \not{p} + m, \qquad \sum_s v^s(p)\bar{v}^s(p) = \not{p} - m. \tag{A.22}$$

For polarized cross sections one can either resort to the explicit formulae (A.19) or insert the projection matrices

$$\left(\frac{1+\gamma^5}{2}\right), \qquad \left(\frac{1-\gamma^5}{2}\right), \tag{A.23}$$

which project onto right- and left-handed spinors, respectively. Again, for antifermions, the helicity of the spinor is opposite to the physical helicity of the particle.

Many other identities involving Dirac spinors and matrices can be found in Chapter 3.

Polarization vectors for photons and other gauge bosons are conventionally normalized to unity. For massless bosons the polarization must be transverse:

$$\epsilon^\mu = (0, \boldsymbol{\epsilon}), \qquad \text{where } \mathbf{p} \cdot \boldsymbol{\epsilon} = 0. \tag{A.24}$$

If $\mathbf{p}$ is in the $+z$ direction, the polarization vectors are

$$\epsilon^\mu = \frac{1}{\sqrt{2}}(0, 1, i, 0), \qquad \epsilon^\mu = \frac{1}{\sqrt{2}}(0, 1, -i, 0), \tag{A.25}$$

for right- and left-handed helicities, respectively.

In computing unpolarized cross sections involving photons, one can replace

$$\sum_{\text{polarizations}} \epsilon^*_\mu \epsilon_\nu \longrightarrow -g_{\mu\nu}, \tag{A.26}$$

by virtue of the Ward identity. In the case of massless non-Abelian gauge bosons, one must also sum over the emission of ghosts, as discussed in Section 16.3. In the massive case, one must in addition include the emission of Goldstone bosons, as discussed in Section 21.1.

A.3 Numerator Algebra

Traces of γ matrices can be evaluated as follows:

$$\begin{aligned}
\operatorname{tr}(\mathbf{1}) &= 4 \\
\operatorname{tr}(\text{any odd \# of } \gamma\text{'s}) &= 0 \\
\operatorname{tr}(\gamma^\mu\gamma^\nu) &= 4g^{\mu\nu} \\
\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma) &= 4(g^{\mu\nu}g^{\rho\sigma} - g^{\mu\rho}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\rho}) \\
\operatorname{tr}(\gamma^5) &= 0 \\
\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^5) &= 0 \\
\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma^5) &= -4i\epsilon^{\mu\nu\rho\sigma}
\end{aligned} \tag{A.27}$$

Another identity allows one to reverse the order of γ matrices inside a trace:

$$\operatorname{tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\cdots) = \operatorname{tr}(\cdots\gamma^\sigma\gamma^\rho\gamma^\nu\gamma^\mu). \tag{A.28}$$

Contractions of γ matrices with each other simplify to:

$$\begin{aligned}
\gamma^\mu\gamma_\mu &= 4 \\
\gamma^\mu\gamma^\nu\gamma_\mu &= -2\gamma^\nu \\
\gamma^\mu\gamma^\nu\gamma^\rho\gamma_\mu &= 4g^{\nu\rho} \\
\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma_\mu &= -2\gamma^\sigma\gamma^\rho\gamma^\nu
\end{aligned} \tag{A.29}$$

(These identities apply in four dimensions only; see the following section.) Contractions of the ϵ symbol can also be simplified:

$$\begin{aligned}
\epsilon^{\alpha\beta\gamma\delta}\epsilon_{\alpha\beta\gamma\delta} &= -24 \\
\epsilon^{\alpha\beta\gamma\mu}\epsilon_{\alpha\beta\gamma\nu} &= -6\delta^\mu{}_\nu \\
\epsilon^{\alpha\beta\mu\nu}\epsilon_{\alpha\beta\rho\sigma} &= -2\big(\delta^\mu{}_\rho\delta^\nu{}_\sigma - \delta^\mu{}_\sigma\delta^\nu{}_\rho\big)
\end{aligned} \tag{A.30}$$

In some calculations, it is useful to rearrange products of fermion bilinears by means of *Fierz identities.* Let $u_1, \ldots, u_4$ be Dirac spinors, and let $u_{iL} = \frac{1}{2}(1-\gamma^5)u_i$ be the left-handed projection. Then the most important Fierz rearrangement formula is

$$(\bar{u}_{1L}\gamma^\mu u_{2L})(\bar{u}_{3L}\gamma_\mu u_{4L}) = -(\bar{u}_{1L}\gamma^\mu u_{4L})(\bar{u}_{3L}\gamma_\mu u_{2L}). \tag{A.31}$$

Additional formulae can be generated by the use of the following identities for the 2×2 blocks of Dirac matrices:

$$(\sigma^\mu)_{\alpha\beta}(\sigma_\mu)_{\gamma\delta} = 2\epsilon_{\alpha\gamma}\epsilon_{\beta\delta}; \qquad (\bar{\sigma}^\mu)_{\alpha\beta}(\bar{\sigma}_\mu)_{\gamma\delta} = 2\epsilon_{\alpha\gamma}\epsilon_{\beta\delta}. \tag{A.32}$$

In non-Abelian gauge theories, the Feynman rules involve gauge group matrices t^a that form a representation r of a Lie algebra G. The symbol G also denotes the adjoint representation of the algebra. The matrices t^a obey

$$[t^a, t^b] = if^{abc}t^c, \tag{A.33}$$

where the structure constants f^{abc} are totally antisymmetric. The invariants $C(r)$ and $C_2(r)$ of the representation r are defined by

$$\mathrm{tr}[t^a t^b] = C(r)\delta^{ab}, \qquad t^a t^a = C_2(r)\cdot \mathbf{1}. \tag{A.34}$$

These are related by

$$C(r) = \frac{d(r)}{d(G)}C_2(r), \tag{A.35}$$

where $d(r)$ is the dimension of the representation. Traces and contractions of the t^a can be evaluated using the above identities and their consequences:

$$\begin{aligned} t^a t^b t^a &= [C_2(r) - \tfrac{1}{2}C_2(G)]t^b \\ f^{acd}f^{bcd} &= C_2(G)\delta^{ab} \\ f^{abc}t^b t^c &= \tfrac{1}{2}iC_2(G)t^a \end{aligned} \tag{A.36}$$

For $SU(N)$ groups, the fundamental representation is denoted by N, and we have

$$C(N) = \frac{1}{2}, \qquad C_2(N) = \frac{N^2-1}{2N}, \qquad C(G) = C_2(G) = N. \tag{A.37}$$

The following relation, satisfied by the matrices of the fundamental representation of $SU(N)$, is also very helpful:

$$(t^a)_{ij}(t^a)_{k\ell} = \frac{1}{2}\Big(\delta_{i\ell}\delta_{kj} - \frac{1}{N}\delta_{ij}\delta_{k\ell}\Big). \tag{A.38}$$

A.4 Loop Integrals and Dimensional Regularization

To combine propagator denominators, introduce integrals over Feynman parameters:

$$\frac{1}{A_1A_2\cdots A_n} = \int_0^1 dx_1\cdots dx_n\,\delta(\textstyle\sum x_i - 1)\,\dfrac{(n-1)!}{\big[x_1A_1 + x_2A_2 + \cdots x_nA_n\big]^n} \tag{A.39}$$

In the case of only two denominator factors, this formula reduces to

$$\frac{1}{AB} = \int_0^1 dx\,\frac{1}{\big[xA + (1-x)B\big]^2}. \tag{A.40}$$

A more general formula in which the A_i are raised to arbitrary powers is given in Eq. (6.42).

Once this is done, the bracketed quantity in the denominator will be a quadratic function of the integration variables p_i^μ. Next, complete the square and shift the integration variables to absorb the terms linear in p_i^μ. For a one-loop integral, there is a single integration momentum p^μ, which is shifted to a momentum variable ℓ^μ. After this shift, the denominator takes the form

$(\ell^2 - \Delta)^n$. In the numerator, terms with an odd number of powers of ℓ vanish by symmetric integration. Symmetry also allows one to replace

$$\ell^\mu \ell^\nu \to \frac{1}{d}\ell^2 g^{\mu\nu}, \tag{A.41}$$

$$\ell^\mu \ell^\nu \ell^\rho \ell^\sigma \to \frac{1}{d(d+2)}(\ell^2)^2\big(g^{\mu\nu}g^{\rho\sigma} + g^{\mu\rho}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\rho}\big). \tag{A.42}$$

(Here d is the spacetime dimension.) The integral is most conveniently evaluated after Wick-rotating to Euclidean space, with the substitution

$$\ell^0 = i\ell^0_E, \qquad \ell^2 = -\ell^2_E. \tag{A.43}$$

Alternatively, one can use the following table of d-dimensional integrals in Minkowski space:

$$\int \frac{d^d\ell}{(2\pi)^d}\frac{1}{(\ell^2-\Delta)^n} = \frac{(-1)^n i}{(4\pi)^{d/2}}\frac{\Gamma(n-\frac{d}{2})}{\Gamma(n)}\left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}} \tag{A.44}$$

$$\int \frac{d^d\ell}{(2\pi)^d}\frac{\ell^2}{(\ell^2-\Delta)^n} = \frac{(-1)^{n-1} i}{(4\pi)^{d/2}}\frac{d}{2}\frac{\Gamma(n-\frac{d}{2}-1)}{\Gamma(n)}\left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}-1} \tag{A.45}$$

$$\int \frac{d^d\ell}{(2\pi)^d}\frac{\ell^\mu\ell^\nu}{(\ell^2-\Delta)^n} = \frac{(-1)^{n-1} i}{(4\pi)^{d/2}}\frac{g^{\mu\nu}}{2}\frac{\Gamma(n-\frac{d}{2}-1)}{\Gamma(n)}\left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}-1} \tag{A.46}$$

$$\int \frac{d^d\ell}{(2\pi)^d}\frac{(\ell^2)^2}{(\ell^2-\Delta)^n} = \frac{(-1)^n i}{(4\pi)^{d/2}}\frac{d(d+2)}{4}\frac{\Gamma(n-\frac{d}{2}-2)}{\Gamma(n)}\left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}-2} \tag{A.47}$$

$$\int \frac{d^d\ell}{(2\pi)^d}\frac{\ell^\mu\ell^\nu\ell^\rho\ell^\sigma}{(\ell^2-\Delta)^n} = \frac{(-1)^n i}{(4\pi)^{d/2}}\frac{\Gamma(n-\frac{d}{2}-2)}{\Gamma(n)}\left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}-2} \times \frac{1}{4}\big(g^{\mu\nu}g^{\rho\sigma} + g^{\mu\rho}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\rho}\big) \tag{A.48}$$

If the integral converges, one can set $d = 4$ from the start. If the integral diverges, the behavior near $d = 4$ can be extracted by expanding

$$\left(\frac{1}{\Delta}\right)^{2-\frac{d}{2}} = 1 - (2-\tfrac{d}{2})\log\Delta + \cdots. \tag{A.49}$$

One also needs the expansion of $\Gamma(x)$ near its poles:

$$\Gamma(x) = \frac{1}{x} - \gamma + \mathcal{O}(x) \tag{A.50}$$

near $x = 0$, and

$$\Gamma(x) = \frac{(-1)^n}{n!}\left(\frac{1}{x+n} - \gamma + 1 + \cdots + \frac{1}{n} + \mathcal{O}(x+n)\right) \tag{A.51}$$

near $x = -n$. Here γ is the Euler-Mascheroni constant, $\gamma \approx 0.5772$. The following combination of terms often appears in calculations:

$$\frac{\Gamma(2-\frac{d}{2})}{(4\pi)^{d/2}}\left(\frac{1}{\Delta}\right)^{2-\frac{d}{2}} = \frac{1}{(4\pi)^2}\left(\frac{2}{\epsilon} - \log\Delta - \gamma + \log(4\pi) + \mathcal{O}(\epsilon)\right), \tag{A.52}$$

with $\epsilon = 4 - d$.

Notice that Δ is positive if it is a combination of masses and *spacelike* momentum invariants. If Δ contains timelike momenta, it may become negative. Then these integrals acquire imaginary parts, which give the discontinuities of S-matrix elements. To compute the S-matrix in a physical region, choose the correct branch of the function by the prescription

$$\left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}} \rightarrow \left(\frac{1}{\Delta - i\epsilon}\right)^{n-\frac{d}{2}}, \tag{A.53}$$

where $-i\epsilon$ (not to be confused with ϵ in the previous paragraph!) gives a tiny negative imaginary part.

Traces in Eq. (A.27) that do not involve γ^5 are independent of dimensionality. However, since

$$g^{\mu\nu}g_{\mu\nu} = \delta^{\mu}{}_{\mu} = d \tag{A.54}$$

in d dimensions, the contraction identities (A.29) are modified:

$$\begin{aligned} \gamma^{\mu}\gamma_{\mu} &= d \\ \gamma^{\mu}\gamma^{\nu}\gamma_{\mu} &= -(d-2)\gamma^{\nu} \\ \gamma^{\mu}\gamma^{\nu}\gamma^{\rho}\gamma_{\mu} &= 4g^{\nu\rho} - (4-d)\gamma^{\nu}\gamma^{\rho} \\ \gamma^{\mu}\gamma^{\nu}\gamma^{\rho}\gamma^{\sigma}\gamma_{\mu} &= -2\gamma^{\sigma}\gamma^{\rho}\gamma^{\nu} + (4-d)\gamma^{\nu}\gamma^{\rho}\gamma^{\sigma} \end{aligned} \tag{A.55}$$

A.5 Cross Sections and Decay Rates

Once the squared matrix element for a scattering process is known, the differential cross section is given by

$$\begin{aligned} d\sigma = &\frac{1}{2E_{\mathcal{A}}2E_{\mathcal{B}}\,|v_{\mathcal{A}}-v_{\mathcal{B}}|}\left(\prod_f \frac{d^3p_f}{(2\pi)^3}\frac{1}{2E_f}\right) \\ &\times \left|\mathcal{M}(p_{\mathcal{A}}, p_{\mathcal{B}} \rightarrow \{p_f\})\right|^2 (2\pi)^4\delta^{(4)}(p_{\mathcal{A}}+p_{\mathcal{B}} - \textstyle\sum p_f). \end{aligned} \tag{A.56}$$

The differential decay rate of an unstable particle to a given final state is

$$d\Gamma = \frac{1}{2m_{\mathcal{A}}}\left(\prod_f \frac{d^3p_f}{(2\pi)^3}\frac{1}{2E_f}\right)\left|\mathcal{M}(m_{\mathcal{A}} \rightarrow \{p_f\})\right|^2 (2\pi)^4\delta^{(4)}(p_{\mathcal{A}} - \textstyle\sum p_f). \tag{A.57}$$

For the special case of a two-particle final state, the Lorentz-invariant phase space takes the simple form

$$\left(\prod_f \int \frac{d^3p_f}{(2\pi)^3}\frac{1}{2E_f}\right)(2\pi)^4\delta^{(4)}(\textstyle\sum p_i - \sum p_f) = \displaystyle\int \frac{d\Omega_{\rm cm}}{4\pi}\,\frac{1}{8\pi}\left(\frac{2|\mathbf{p}|}{E_{\rm cm}}\right), \tag{A.58}$$

where $|\mathbf{p}|$ is the magnitude of the 3-momentum of either particle in the center-of-mass frame.

A.6 Physical Constants and Conversion Factors

Precisely known physical constants:

$$
\begin{aligned}
c &= 2.998 \times 10^{10} \text{ cm/s} \\
\hbar &= 6.582 \times 10^{-22} \text{ MeV s} \\
e &= -1.602 \times 10^{-19} \text{ C} \\
\alpha &= \frac{e^2}{4\pi\hbar c} = \frac{1}{137.04} = 0.00730 \\
\frac{G_F}{(\hbar c)^3} &= 1.166 \times 10^{-5} \text{ GeV}^{-2}
\end{aligned}
$$

The values of the strong and weak interaction coupling constants depend on the conventions used to define them, as explained in Sections 17.6 and 21.3. For the purpose of estimation, however, one can use the following values:

$$
\begin{aligned}
\alpha_s(10 \text{ GeV}) &= 0.18 \\
\alpha_s(m_Z) &= 0.12 \\
\sin^2\theta_w &= 0.23
\end{aligned}
$$

Particle masses (times c^2):

e :	0.5110 MeV	p :	938.3 MeV
μ :	105.6 MeV	n :	939.6 MeV
τ :	1777 MeV	$\pi^\pm$:	139.6 MeV
$W^\pm$:	80.2 GeV	π^0 :	135.0 MeV
Z^0 :	91.19 GeV		

Useful combinations:

Bohr radius: $$a_0 = \frac{\hbar}{\alpha m_e c} = 5.292 \times 10^{-9} \text{ cm}$$

electron Compton wavelength: $$\lambdabar = \frac{\hbar}{m_e c} = 3.862 \times 10^{-11} \text{ cm}$$

classical electron radius: $$r_e = \frac{\alpha\hbar}{m_e c} = 2.818 \times 10^{-13} \text{ cm}$$

Thomson cross section: $$\sigma_T = \frac{8\pi r_e^2}{3} = 0.6652 \text{ barn}$$

annihilation cross section: $$1 \text{ R} = \frac{4\pi\alpha^2}{3E_{\text{cm}}^2} = \frac{86.8 \text{ nbarn}}{(E_{\text{cm}} \text{ in GeV})^2}$$

Conversion factors:

$$
\begin{aligned}
(1\ \text{GeV})/c^2 &= 1.783 \times 10^{-24}\ \text{g} \\
(1\ \text{GeV})^{-1}(\hbar c) &= 0.1973 \times 10^{-13}\ \text{cm} = 0.1973\ \text{fm}; \\
(1\ \text{GeV})^{-2}(\hbar c)^2 &= 0.3894 \times 10^{-27}\ \text{cm}^2 = 0.3894\ \text{mbarn} \\
1\ \text{barn} &= 10^{-24}\ \text{cm}^2 \\
(1\ \text{volt/meter})(e\hbar c) &= 1.973 \times 10^{-25}\ \text{GeV}^2 \\
(1\ \text{tesla})(e\hbar c^2) &= 5.916 \times 10^{-17}\ \text{GeV}^2
\end{aligned}
$$

A complete, up-to-date tabulation of the fundamental constants and the properties of elementary particles is given in the *Review of Particle Properties*, which can be found in a recent issue of either *Physical Review* **D** or *Physics Letters* **B**. The most recent *Review* as of this writing is published in *Physical Review* **D50**, 1173 (1994).

Bibliography

Mathematical Background and Reference

Cahn, Robert N., *Semi-Simple Lie Algebras and Their Representations*, Benjamin/Cummings, Menlo Park, California, 1984. Written by a physicist, for physicists.

Carrier, George F., Krook, Max, and Pearson, Carl E., *Functions of a Complex Variable*, McGraw-Hill, New York, 1966. An excellent practical introduction to methods of complex variables and contour integration.

Gradshteyn, I. S. and Ryzhik, I. M., *Table of Integrals, Series, and Products* (trans. and ed. by Alan Jeffrey), Academic Press, Orlando, Florida, 1980.

Physics Background

Baym, Gordon, *Lectures on Quantum Mechanics*, Benjamin/Cummings, Menlo Park, California, 1969. A concise, informal text that is especially rich in nontrivial applications.

Fetter, Alexander L. and Walecka, John Dirk, *Theoretical Mechanics of Particles and Continua*, McGraw-Hill, New York, 1980. Includes several chapters on continuum mechanics and useful appendices of mathematical reference material.

Feynman, Richard P., *QED: The Strange Theory of Light and Matter*, Princeton University Press, Princeton, New Jersey, 1985. A transcription of four lectures given to a general audience, presenting Feynman's approach to quantum mechanics, including Feynman diagrams. Highly recommended.

Feynman, Richard P. and Hibbs, A. R., *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, 1965. An introduction to the use of path integrals in nonrelativistic quantum mechanics.

Goldstein, Herbert, *Classical Mechanics* (second edition), Addison-Wesley, Reading, Massachusetts, 1980. Chapter 12 introduces classical relativistic field theory.

Jackson, J. D., *Classical Electrodynamics* (second edition), Wiley, New York, 1975.

Landau, L. D. and Lifshitz, E. M., *The Classical Theory of Fields* (fourth revised English edition, trans. Morton Hamermesh), Pergamon, Oxford, 1975. Contains a succinct development of electromagnetic theory from the Lagrangian viewpoint.

Landau, L. D. and Lifshitz, E. M., *Statistical Physics* (third edition, Part 1, trans. J. B. Sykes and M. J. Kearsley), Pergamon Press, 1980. An insightful if concise textbook of statistical mechanics, containing the original pedagogical exposition of Landau's theory of phase transitions.

Reichl, L. E., *A Modern Course in Statistical Physics*, University of Texas Press, Austin, 1980. A complete textbook of statistical mechanics.

Schiff, Leonard I., *Quantum Mechanics* (third edition), McGraw-Hill, New York, 1968.

Shankar, Ramamurti, *Principles of Quantum Mechanics*, Plenum, New York, 1980. A very clear presentation of the basic theory.

Taylor, Edwin F. and Wheeler, John Archibald, *Spacetime Physics* (second edition), Freeman, New York, 1992. An elementary but insightful introduction to special relativity.

Taylor, John R., *Scattering Theory*, Robert E. Krieger, Malabar, Florida, 1983 (reprint of original edition published by Wiley, New York, 1972). A very clear development of scattering theory for nonrelativistic quantum mechanics.

Relativistic Quantum Mechanics and Field Theory

Bailin, D. and Love, A., *Introduction to Gauge Field Theory* (revised edition), Institute of Physics Publishing, Bristol, 1993. Develops the theory entirely from the path integral viewpoint.

Balian, Roger and Zinn-Justin, Jean (eds.), *Methods in Field Theory*, North-Holland, Amsterdam, 1976. Proceedings of the 1975 Les Houches Summer School in Theoretical Physics, including lectures on functional methods, renormalization, and gauge theories.

Berestetskii, V. B., Lifshitz, E. M., and Pitaevskii, L. P., *Quantum Electrodynamics* (second edition, trans. J. B. Sykes and J. S. Bell), Pergamon, Oxford, 1982. An excellent reference for QED applications.

Bjorken, James D. and Drell, Sidney D., *Relativistic Quantum Mechanics*, McGraw-Hill, New York, 1964. Develops Feynman diagrams using intuitive arguments, without using fields.

Bjorken, James D. and Drell, Sidney D., *Relativistic Quantum Fields*, McGraw-Hill, New York, 1965. Redevelops Feynman diagrams from the field viewpoint, using canonical quantization.

Brown, Lowell S., *Quantum Field Theory*, Cambridge University Press, New York, 1992. A careful treatment of the foundations of quantum field theory and its application to scattering processes.

Coleman, Sidney, *Aspects of Symmetry*, Cambridge University Press, Cambridge, 1985. Informal lectures on a number of topics involving gauge theories and symmetry, given between 1966 and 1979.

Collins, John, *Renormalization*, Cambridge University Press, Cambridge, 1984. A careful development of the technical machinery needed for all-orders proofs of renormalizability, operator product expansion, and factorization theorems.

Deser, Stanley, Grisaru, Marc, and Pendleton, Hugh, *Lectures on Elementary Particles and Quantum Field Theory*, MIT Press, Cambridge, 1970, vol. 1. Four extremely useful summer school lectures.

Gross, Franz, *Relativistic Quantum Mechanics and Field Theory*, Wiley, New York, 1993. Includes a number of topics in "advanced quantum mechanics", and an introductory chapter on bound states.

Itzykson, Claude and Zuber, Jean-Bernard, *Quantum Field Theory*, McGraw-Hill, New York, 1980. A comprehensive textbook.

Jauch, J. M. and Rohrlich, F., *The Theory of Photons and Electrons* (second edition), Springer-Verlag, Berlin, 1976. An authoritative monograph on QED.

Kaku, Michio, *Quantum Field Theory: A Modern Introduction*, Oxford University Press, New York, 1993. Contains brief introductory chapters on a number of advanced topics.

Kinoshita, T., ed., *Quantum Electrodynamics*, World Scientific, Singapore, 1990. A collection of review articles on precision tests of QED.

Mandl, F. and Shaw, G., *Quantum Field Theory* (revised edition), Wiley, New York, 1993. The easiest book on field theory; introduces QED and some electroweak theory using canonical quantization.

Ramond, Pierre, *Field Theory: A Modern Primer* (second edition), Addison-Wesley, Redwood City, California, 1989. Contains very nice treatments of the Lorentz group, path integrals, ϕ^4 theory, and quantization of gauge theories.

Ryder, Lewis H., *Quantum Field Theory*, Cambridge University Press, Cambridge, 1985. A concise treatment of the more formal aspects of the subject.

Sakurai, J. J., *Advanced Quantum Mechanics*, Addison-Wesley, Reading, 1967. Develops Feynman diagrams without using fields.

Schweber, Silvan S., *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*, Princeton University Press, Princeton, 1994. An excellent history of the subject up to about 1950.

Schwinger, Julian (ed.), *Selected Papers on Quantum Electrodynamics*, Dover, New York, 1958. Reprints of important papers written between 1927 and 1953.

Sterman, George, *Introduction to Quantum Field Theory*, Cambridge University Press, Cambridge, 1993. An introductory textbook with special emphasis on the essentials of perturbative QCD.

Elementary Particle Physics

Aitchison, Ian J. R. and Hey, Anthony J. G., *Gauge Theories in Particle Physics* (second edition), Adam Hilger, Bristol, 1989. An elementary introduction to gauge theories, concentrating mostly on tree-level processes.

Barger, Vernon and Phillips, Roger J. N., *Collider Physics*, Addison-Wesley, Menlo Park, California, 1987. Basic discussion of the application of QCD to high-energy collider phenomenology.

Cahn, Robert N. and Goldhaber, Gerson, *The Experimental Foundations of Particle Physics*, Cambridge University Press, Cambridge, 1989. Reprints of many original papers, supplemented by introductory overviews, additional references, and exercises. Highly recommended.

Cheng, Ta-Pei and Li, Ling-Fong, *Gauge Theory of Elementary Particle Physics*, Oxford University Press, New York, 1984. An advanced, authoritative monograph.

Commins, Eugene D. and Bucksbaum, Philip H., *Weak Interactions of Leptons and Quarks*, Cambridge University Press, Cambridge, 1983. A thorough review of both theory and experiment.

Field, Richard D., *Applications of Perturbative QCD*, Benjamin/Cummings, Menlo Park, 1989. A useful description of the techniques needed for QCD calculations beyond the leading order.

Georgi, Howard, *Weak Interactions and Modern Particle Theory*, Benjamin/Cummings, Menlo Park, California, 1984. Advanced, insightful treatment of selected topics.

Griffiths, David, *Introduction to Elementary Particles*, Wiley, New York, 1987. A good undergraduate-level survey.

Halzen, Francis and Martin, Alan D., *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, Wiley, New York, 1984. Uses Feynman diagrams throughout, and concentrates on gauge theories.

Perkins, Donald H., *Introduction to High Energy Physics* (third edition), Addison-Wesley, Menlo Park, California, 1987. A good introduction to phenomena, with relatively little emphasis on Feynman diagrams and gauge theories.

Quigg, Chris, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*, Benjamin/Cummings, Menlo Park, California, 1983. A very nice overview of gauge theories and their experimental tests.

Ross, Graham G., *Grand Unified Theories*, Benjamin/Cummings, Menlo Park, California, 1984. A clear introduction to gauge theories that unify the interactions of particle physics.

Taylor, J. C., *Gauge Theories of Weak Interactions*, Cambridge University Press, Cambridge, 1976. A concise treatment of the standard model and related theoretical issues.

Condensed Matter Physics

Abrikosov, A. A., Gorkov, L. P., and Dzyaloshinskii, I. E., *Quantum Field Theoretical Methods in Statistical Physics* (second edition), Pergamon, Oxford, 1965. A classic, but very terse, exposition of the application of Feynman diagrams to condensed matter problems.

Anderson, P. W., *Basic Notions of Condensed Matter Physics*, Benjamin/Cummings, Menlo Park, California, 1984. An informal overview of the concepts of broken symmetry and renormalization as applied to condensed matter systems.

Fetter, Alexander L. and Walecka, John Dirk, *Quantum Theory of Many-Particle Systems*, McGraw-Hill, New York, 1971. A straightfoward introduction to the use of Feynman diagrams in nuclear and condensed matter physics.

Ma, Shang-Keng, *Modern Theory of Critical Phenomena*, Benjamin/Cummings, 1976. An introduction to the use of renormalization group methods in the theory of critical phenomena.

Mattuck, Richard D., *A Guide to Feynman Diagrams in the Many-Body Problem*, McGraw-Hill, New York, 1967. A clear and easy introduction to the use of Feynman diagrams in solid state physics.

Parisi, Giorgio, *Statistical Field Theory*, Benjamin/Cummings, 1988. Far-ranging applications of ideas from quantum field theory to problems in statistical mechanics.

Stanley, H. Eugene, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford, 1971.

Zinn-Justin, Jean, *Quantum Field Theory and Critical Phenomena* (third edition), Oxford University Press, Oxford, 1996. A treatise on the application of renormalization theory to the study of critical phenomena.

Corrections to This Book

A list of misprints and corrections to this book is posted on the World-Wide Web at the URL 'http://www.slac.stanford.edu/~mpeskin/QFT.html', or can be obtained by writing to the authors. We would be grateful if you would report additional errors in the book, or send other comments, to mpeskin@slac.stanford.edu or to dschroeder@cc.weber.edu.

Index